

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 04:54:54 ; Search time 2762.24 Seconds
(without alignments)
65.021 Million cell updates/sec

Title: US-09-142-095-4

Perfect score: 19

Sequence: 1 ttgtctctgcagaggtt 19

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1028115 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est18:*
118: gb_est19:*
119: gb_est20:*
120: gb_est21:*
121: gb_est22:*
122: gb_est23:*
123: gb_est24:*
124: gb_est25:*
125: gb_est26:*
126: gb_est27:*
127: gb_est28:*
128: gb_est29:*
129: gb_est30:*
130: gb_est31:*
131: gb_est32:*
132: gb_est33:*
133: gb_est34:*
134: gb_est35:*
135: gb_est36:*
136: gb_est37:*
137: gb_est38:*
138: gb_est39:*
139: gb_est40:*
140: gb_est41:*
141: gb_est42:*
142: gb_est43:*
143: gb_est44:*
144: gb_est45:*
145: gb_est46:*
146: gb_est47:*
147: gb_est48:*
148: gb_est49:*
149: gb_est50:*
150: gb_est51:*
151: gb_est52:*
152: gb_est53:*
153: gb_est54:*
154: gb_est55:*
155: gb_est56:*
156: gb_est57:*
157: gb_est58:*
158: gb_est59:*
159: gb_est60:*
160: gb_est61:*
161: gb_est62:*
162: gb_est63:*
163: gb_est64:*
164: gb_est65:*
165: gb_est66:*
166: gb_est67:*
167: gb_est68:*
168: gb_est69:*
169: gb_est70:*
170: gb_est71:*
171: gb_est72:*
172: gb_est73:*
173: gb_est74:*
174: gb_est75:*
175: gb_est76:*
176: gb_est77:*
177: gb_est78:*
178: gb_est79:*
179: gb_est80:*
180: gb_est81:*
181: gb_est82:*
182: gb_est83:*
183: gb_est84:*
184: gb_est85:*
185: gb_est86:*
186: gb_est87:*
187: gb_est88:*
188: gb_est89:*
189: gb_est90:*

190: gb_est91:*
191: gb_est92:*
192: gb_est93:*
193: gb_est94:*
194: gb_est95:*
195: gb_est96:*
196: gb_est97:*
197: gb_est98:*
198: gb_est99:*
199: gb_est100:*
200: gb_est101:*
201: gb_est102:*
202: gb_est103:*
203: gb_est104:*
204: gb_est105:*
205: gb_est106:*
206: gb_est107:*
207: gb_est108:*
208: gb_est109:*
209: gb_est110:*
210: gb_est111:*
211: gb_est112:*
212: gb_est113:*
213: gb_est114:*
214: gb_est115:*
215: gb_est116:*
216: gb_est117:*
217: gb_est118:*
218: gb_est119:*
219: gb_est120:*
220: gb_est121:*
221: gb_est122:*
222: gb_est123:*
223: gb_est124:*
224: gb_est125:*
225: gb_est126:*
226: gb_est127:*
227: gb_est128:*
228: gb_est129:*
229: gb_est130:*
230: gb_est131:*
231: gb_est132:*
232: gb_est133:*
233: gb_est134:*
234: gb_est135:*
235: gb_est136:*
236: gb_est137:*
237: gb_est138:*
238: gb_est139:*
239: gb_est140:*
240: gb_est141:*
241: gb_est142:*
242: gb_est143:*
243: gb_est144:*
244: gb_est145:*
245: gb_est146:*
246: gb_est147:*
247: gb_est148:*
248: gb_est149:*
249: gb_est150:*
250: gb_est151:*
251: gb_est152:*
252: gb_est153:*
253: gb_est154:*
254: gb_est155:*
255: gb_est156:*
256: gb_est157:*
257: gb_est158:*
258: gb_est159:*
259: gb_est160:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.4	91.6	522	150	BF545097
2	17.4	91.6	608	230	AO540572
3	17.4	89.5	367	115	AM429811
4	17.4	89.5	936	175	BG297415
5	16.4	86.3	356	223	AO68059
6	16.4	86.3	530	233	AO799987
7	16.4	86.3	574	104	A1979827
8	16.4	86.3	616	247	A2664475
9	16.4	86.3	680	141	BE901069
10	16.4	86.3	978	173	BG109825
11	16.4	86.3	206	17	AT203923
12	16.4	86.3	338	8	AA504455
13	16.4	86.3	338	8	AA504455
14	16.4	86.3	338	8	AA504455
15	16.4	86.3	338	8	AA504455
16	16.4	86.3	338	8	AA504455
17	16.4	86.3	338	8	AA504455
18	16.4	86.3	338	8	AA504455
19	16.4	86.3	338	8	AA504455
20	16.4	86.3	338	8	AA504455
21	16.4	86.3	338	8	AA504455
22	16.4	86.3	338	8	AA504455
23	16.4	86.3	338	8	AA504455
24	16.4	86.3	338	8	AA504455
25	16.4	86.3	338	8	AA504455
26	16.4	86.3	338	8	AA504455
27	16.4	86.3	338	8	AA504455
28	16.4	86.3	338	8	AA504455
29	16.4	86.3	338	8	AA504455
30	16.4	86.3	338	8	AA504455
31	16.4	86.3	338	8	AA504455
32	16.4	86.3	338	8	AA504455
33	16.4	86.3	338	8	AA504455
34	16.4	86.3	338	8	AA504455
35	16.4	86.3	338	8	AA504455
36	16.4	86.3	338	8	AA504455
37	16.4	86.3	338	8	AA504455
38	16.4	86.3	338	8	AA504455
39	16.4	86.3	338	8	AA504455
40	16.4	86.3	338	8	AA504455
41	16.4	86.3	338	8	AA504455
42	16.4	86.3	338	8	AA504455
43	16.4	86.3	338	8	AA504455
44	16.4	86.3	338	8	AA504455
45	16.4	86.3	338	8	AA504455

ALIGNMENTS

RESULT 1
LOCUS BF545097
DEFINITION BF545097 522 bp mRNA
ACCESSION BF545097
VERSION BF545097.1 GI:11636204
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (Bases 1 to 522)
AUTHORS Bonaldo, M.F., Lennon, G., and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL MEDLINE COMMENT

Genome Res. 6 (9), 791-806 (1996)
9704447
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
CDNA Library Preparation: M.B. Soares Lab Clone Distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
ILMIL (info@imgc.llnl.gov). IMAGE ID- 1789283
Seq primer: M3 forward.

FEATURES

source
1. 522
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C2p-gt-e-07-0-UI"
/clone="UI-R-C2p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pTZ19-pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-C2p library is a subtracted library derived from the UI-R-C1 library, which is a subtracted library derived from the UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996).
BASE COUNT 156 a 117 c 104 g 145 t
ORIGIN

Query Match 91.6%; Score 17.4; DB 150; Length 522;
Best Local Similarity 94.7%; Pred. No. 2, 1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ttgtctctccagaggtt 19
DB 128 TTGCTCTCTCCAGAGGTT 146

RESULT 2
LOCUS AO540572/c
DEFINITION AO540572 608 bp DNA
ACCESSION AO540572
VERSION AO540572.1 GI:4871102
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (Bases 1 to 608)
AUTHORS Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P., and Venter

TITLE
J.C.
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL
Unpublished (1997)
COMMENT
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: dbeetlgr.org

Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: 17
Class: BAC ends.

FEATURES
source
Location/Qualifiers

1..608
/organism="Homo sapiens"
/db_xref="GDB:7637206"
/db_xref="taxon:9606"
/clone="RPCI-11-358E23"
/clone_11b="RPCI-11"
/sex="Male"
/cell_type="lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"
BASE COUNT 259 a 83 c 93 g 173 t
ORIGIN

Query Match 91.6%; Score 17.4; DB 230; Length 608;
Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ttgtctctgcagaggtt 19
|||||
Db 138 TTGGCTCTCTCCAGAGTT 120

RESULT 3
AM429811 367 bp mRNA EST 09-JUL-2000
LOCUS
68306 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
DEFINITION
AM429811
ACCESSION
AM429811.1 GI:6961118
KEYWORDS
EST.

SOURCE
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
1 (bases 1 to 367)
Fahnenkrug, S.C., Fleking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.W.
and Keeler, J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL

JOURNAL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross-match with the -mhscore 20
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCATGACAT
BACKWARD: GTTTCACAGTCAGCAGC

FEATURES
source
Location/Qualifiers

1..367
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_11b="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPOR6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 91 a 100 c 115 g 60 t 1 others
ORIGIN

Query Match 89.5%; Score 17; DB 115; Length 367;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcctctgcagaggtt 19
|||||
Db 256 TGCTCTGCGCAGAGTT 272

RESULT 4
BG297415 936 bp mRNA EST 21-FEB-2001
LOCUS
60295559P1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4507022 5',
DEFINITION
mRNA sequence.
ACCESSION
BG297415
VERSION
BG297415.1 GI:13061044
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 936)
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: gcraps-f@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LHAM0383 row: f column: 15
High quality sequence stop: 604.

FEATURES
source
Location/Qualifiers

1..936
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4507022"
/clone_11b="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPOR6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC library."

BASE COUNT 367 a 218 c 259 g 92 t
ORIGIN

Query Match 89.5%; Score 17; DB 175; Length 936;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:23:03 ; Search time 247.87 Seconds

(without alignments)
53,197 Million cell updates/sec

Title: US-09-142-095-2

Sequence: 1 ccaatggatcaacacgtatct 21

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

N_Geneseq_0601:*

- 1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	21	100.0	21	AAV79542
2	21	100.0	21	AAZ45110
3	21	100.0	1167	UDP-glucuronosyltr
4	21	100.0	2351	UGT1A Exon 1 from
5	15.8	75.2	504	HUG-Brl. Homo sap
6	15.8	75.2	697	Streptococcus pyog
7	15.8	75.2	1073	DNA encoding a hum
8	15.8	75.2	4351	Bacillus thuringle
9	15.8	75.2	4843	LRP5 isoform 2 lon
10	15.8	75.2	4915	LRP5 cDNA longest
11	15.8	75.2	5022	LRP5 isoform 3 put
			19	AAV85549
				LRP5 isoform 5 CDN

12	15.8	75.2	5098	19	AAV70396	LRP5 isoform 1 CDN
13	15.8	75.2	5125	19	AAV86019	LRP5 isoform 3 CDN
14	15.8	75.2	5162	19	AAV85550	LRP5 isoform 6 CDN
15	15.8	75.2	5166	19	AAV85551	LRP5 isoform 2 CDN
16	15.8	75.2	5263	19	AAV70400	LRP5 isoform 4 CDN
17	15.4	73.3	184	13	AAQ27833	Sequence which hyb
18	15.4	73.3	1208	21	AAAB7688	Human secreted pro
19	15.2	72.4	82	12	AAQ14774	Caseln signal pept
20	15.2	72.4	274	21	AAA43269	Xenopus secreted e
21	15.2	72.4	642	20	AAV88862	EST clone H215. H
22	15.2	72.4	691	20	AAV39707	Gastric cancer ass
23	15.2	72.4	1210	12	AAQ14051	Renin gene. Synt
24	15.2	72.4	1312	21	AACT7235	Human ORFX ORP2790
25	15.2	72.4	1416	20	AAE17647	Human gene express
26	15.2	72.4	1900	10	AAV92624	Nucleotide sequenc
27	15.2	72.4	7498	21	AAAB3786	Haemophilus paraga
28	15.2	72.4	8930	21	AAV22834	POLYnucleotide seq
29	15.2	72.4	20682	20	AAV20569	Human secreted pro
30	15	71.4	434	21	AAAC05486	Banana fruit ripen
31	15	71.4	715	20	AAV69488	Human secreted exp
32	14.8	70.5	260	21	AAA44168	Human secreted exp
33	14.8	70.5	297	21	AAA15919	Human protein clon
34	14.8	70.5	300	20	AAE13302	Human gene express
35	14.8	70.5	354	20	AAV41509	Human secreted pro
36	14.8	70.5	379	21	AAA41312	Human secreted exp
37	14.8	70.5	392	21	AAV16090	Human colon cancer
38	14.8	70.5	413	20	AAV97873	Human secreted pro
39	14.8	70.5	415	21	AAAC00147	Human secreted pro
40	14.8	70.5	459	21	AAA38007	UCK-2 nucleotide s
41	14.8	70.5	467	21	AAA15929	Human protein clon
42	14.8	70.5	495	20	AAE234052	Human EST DN443509
43	14.8	70.5	495	21	AACT8499	Human EST DN443509
44	14.8	70.5	495	22	AAE93358	CDNA encoding SRT
45	14.8	70.5	500	20	AAV97826	Human secreted pro

ALIGNMENTS

RESULT 1	
AAV79542	AAV79542 standard; DNK: 21 Bp.
XX	AAV79542;
AC	23-JAN-1998 (first entry)
XX	XX
DT	UGT1*1 gene exon 1 upstream PCR primer B.
XX	XX
DE	Uridine diphosphate glucuronosyltransferase gene; UGT;
KM	Gilbert's syndrome; GS; unconjugated hyperbilirubinaemia;
KM	bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;
KM	Drug trial efficiency; screening; PCR primer; ss.
OS	Synthetic.
OS	Homo sapiens.
XX	XX
PN	W09732042-A2.
XX	XX
PD	04-SEP-1997.
XX	XX
FE	03-MAR-1997; 97WO-GB00577.
XX	XX
PR	16-MAR-1996; 96GB-0005598.
PR	01-MAR-1996; 96GB-0004480.
PA	(UVDU-) UNIV DUNDEE.
XX	XX
PI	Burchell B;
XX	XX
DR	WPI; 1997-448702/41.
XX	XX
PT	Improving drug trial efficiency comprises identifying participants

with Gilbert's syndrome - useful as their altered drug metabolism
may hinder result interpretation

Claim 14; Page 12; 31pp; English.

This PCR primer (with primers AAT79541-44) flanks the TATA box sequence upstream of the uridine diphosphate glucuronosyltransferase (UGT) gene 1A1 exon 1 (see AAT79540), and was used to amplify fragments of 253-255 bp. This gene is known to be associated with Gilbert's syndrome (GS). GS is a mild, common form of unconjugated hyperbilirubinemia associated with reduced bilirubin glucuronidation capacity. Analysis of the genetic basis of GS has allowed 2 forms to be identified. One is a mild form associated with a homozygous 2 bp insertion in the TATA sequence upstream of the UGT1A1 exon 1, and the other is a more severe form associated with heterozygosity for a mutation which, when homozygous, causes Crigler-Najjar type 2 disease. The first form is autosomal recessive and the second is inherited dominantly. Patients suffering from GS, which is benign, may have altered metabolism of some drugs, making it difficult to determine if an effect is due to the drug or the syndrome. Drug trial efficiency would be improved if potential participants can be screened for the genetic basis of GS, and eliminated or included on basis of them possessing or not possessing GS.

Sequence 21 BP; 6 A; 6 C; 4 G; 5 T; 0 other;

Query Match 100.0%; Score 21; DB 18; Length 21;

Best Local Similarity 100.0%; Pred. No. 0.11; 0; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ccactcggatcaacagatct 21
1 ccactcggatcaacagatct 21

RESULT 2

AA245110/c

ID AA245110 standard; DNA; 864 BP.

AC AA245110;

DT 28-FEB-2000 (first entry)

DE UDP-glucuronosyltransferase 1 (UGT1) exon 1A nucleotide sequence.

XX uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;

KW glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;

KM unconjugated hyperbilirubinemia; drug metabolism; transgenic animal;

KM pharmacogenetic screening; diagnose; ss.

XX

OS Homo sapiens.

XX

PN MO957322-A2.

XX

PD 11-NOV-1999.

XX

PE 04-MAY-1999; 99MO-US09702.

XX

PR 07-MAY-1998; 98US-0084807.

XX

PA (AXYS-) AXYS PHARM INC.

XX

PI Penny L. Galvin M.

XX

DR WPI: 2000-052981/04.

XX

P-PSDB: AXY57093.

XX

PT New nucleic acid representing polymorphisms in the human uridine

XX

PT diphosphate glucuronosyltransferase gene, used for diagnosis and

XX

PT evaluation of drug metabolism

XX

PS Examples; Page 27-29; 63pp; English.

Nucleotide sequences AA245110-245118 are exons 1A-1J of the human uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a family of enzymes that catalyze the glucuronic acid conjugation of a wide range of endogenous and exogenous substrates including phenols, alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs result in toxic substances being converted to compounds which are more water soluble and are excreted. The invention relates to and identifies UGT1 polymorphisms (AA24504-245041). The polymorphism sequences are useful as probes for detecting UGT1 locus polymorphisms, indicative of altered UGT1 expression or activity. These polymorphisms are associated with Crigler-Najjar and Gilbert syndromes (unconjugated hyperbilirubinemia) and drug metabolism. The genotyping of the UGT1 gene is used to predict the rate of metabolism of UGT1 substrates, possible drug-drug interactions and adverse side effects (i.e. to optimize drug dosage), and to screen for diseases caused by exposure to toxins and to study the effects of polymorphisms on enzymatic activity. The UGT1 sequences, including polymorphisms, can also be used to produce the corresponding protein (or its fragments) or to generate transgenic animals or modified cells e.g. for pharmacogenetic screening.

Sequence 864 BP; 185 A; 236 C; 209 G; 234 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 864;

Best Local Similarity 100.0%; Pred. No. 0.18;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ccactcggatcaacagatct 21
106 CCACCTGGATCAACAGTACTCT 86

RESULT 3

AA033024/c

ID AA033024 standard; DNA; 1167 BP.

AC AA033024;

XX

DT 27-JAN-1993 (first entry)

DE UGT1A Exon 1 from the UGT1 gene locus.

XX

XX UGT1A; UGT1BP; UGT1C; UGT1D; UGT1E; UGT1F; Isozyme: bilirubin;

KW UDP-glucuronosyl transferase; CN; ss.

XX

OS Homo sapiens.

XX

FT Key

FT msc_RNA

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

10-JAN-1992; 92MO-US00282.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PR 10-JAN-1991; 91US-0639453.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 XX
 PI Owens IS, Ritter JK;
 XX
 DR WPI: 1992-284593/34.
 DR P-PSDB; AAR30194.
 XX
 PT Isolated gene locus UGT1, DNA segments and diagnostic probes -
 PT for diagnosing Gilbert's disease and Crigler-Najjar syndrome
 PT types I and II
 PS Disclosure: Fig 1F; 99pp; English.
 XX
 CC The isolated gene locus, UGT1, has a sequence of about 10000 bp
 CC which represent (1) Exon 1, comprising 6 transcriptional units
 CC (UGT1F, E, D, C, BP and A), represented in AA027368 and
 CC AA033020-24 respectively;
 CC (2) Exon 2, represented in AA033025;
 CC (3) Exon 3, represented in AA033026;
 CC (4) Exon 4, represented in AA033026;
 CC (5) Exon 5, represented in AA033027; and
 CC (6) about 69 kb of non-sequenced DNA.
 CC Six unique N-terminal of 286-289 amino acids are encoded by
 CC the six different first exons and identical C-terminal of 246 amino
 CC acids are encoded by the common exons 2-5. The UGT1 gene locus
 CC encodes a family of UDP-glucuronosyl transferase isozymes, two of
 CC which metabolise bilirubin.
 CC Patients having Crigler-Najjar Syndrome (CN) Type I, have a
 CC mutation present in the second common exon.
 CC
 SQ Sequence 1167 BP; 255 A; 259 C; 272 G; 340 T; 41 other;
 XX
 XX

Query Match 100.0%; Score 21; DB 13; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 0.19; 0; Indels 0; Gaps 0;
 Matches 21; Conservative 0; Mismatches 0;

QY 1 cccactggatcacacagatct 21
 Db 286 CCACCTGGATCACACAGATCT 266.

RESULT 4
 AA027369/C
 ID AA027369 standard; cDNA; 2351 BP.
 XX
 AC AA027369;
 XX
 DT 27-JAN-1993 (first entry)
 XX
 DE HUG-Brl.
 XX
 KW Bilirubin: UDP-glucuronosyltransferase; HUGBrl; HUGB-2;
 KW monoglucuronide; diglucuronide; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 16..784
 FT /tag- a
 FT polyA-signal 2330..2335
 FT /tag- b
 FT /number- 1
 FT polyA-signal 2338..2343
 FT /tag- c
 FT /number- 2
 XX
 PN MO9212987-A.
 XX
 PD 06-AUG-1992.
 XX

PF 10-JAN-1992; 92MO-US00282.
 XX
 PR 10-JAN-1991; 91US-0639453.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 XX
 PI Owens IS, Ritter JK;
 XX
 DR WPI: 1992-284593/34.
 DR P-PSDB; AAR26153.
 XX
 PT Isolated gene locus UGT1, DNA segments and diagnostic probes -
 PT for diagnosing Gilbert's disease and Crigler-Najjar syndrome
 PT types I and II
 PS Disclosure: Fig 9A-I; 99pp; English.
 XX
 CC Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
 CC been isolated. They are referred to as HUGBrl (AA027369) and HUGB-2
 CC (AA027370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
 CC upon expression individually in COS-1 cells, encode isoforms that
 CC catalyse the formation of the two bilirubin monoglucuronides and
 CC the diglucuronide.
 CC The cDNAs contain identical 3' ends (1469 bp in length) to each
 CC other and to that of the human phenol transferase cDNA, HUGP1
 CC (Harding et al., Proc. Natl. Acad. Sci. USA 85:8281 (1988)).
 CC In contrast, they have unique 5' ends.
 CC
 SQ Sequence 2351 BP; 602 A; 540 C; 556 G; 653 T; 0 other;
 XX
 XX

Query Match 100.0%; Score 21; DB 13; Length 2351;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccactggatcacacagatct 21
 Db 121 CCACCTGGATCACACAGATCT 101

RESULT 5
 AAD00564
 ID AAD00564 standard; DNA; 504 BP.
 XX
 AC AAD00564;
 XX
 DT 29-AUG-2000 (first entry)
 XX
 DE Streptococcus pyogenes strain KTL3 partial GRAB protein encoding DNA.
 XX
 KW GRAB protein; protein G related alpha2M binding protein; vaccine;
 KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
 KW Immune response; Streptococcus pyogenes infection; ds.
 XX
 OS Streptococcus pyogenes.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..504
 FT /tag- a
 FT /product- "GRAB protein"
 FT /partial
 XX
 PN MO200026240-A2.
 XX
 PD 11-MAY-2000.
 XX
 PF 02-NOV-1999; 99MO-GB03631.
 XX
 PR 02-NOV-1998; 98GB-0023975.
 XX
 PA (ACTI-) ACTINOVA LTD.
 XX
 PD Bjorck LH, Rasmussen M;
 PI

XX NPI, 2000-365572/31.
 DR P-SDB; AAV71046.
 XX New alpha2M binding protein for generating a protective immune response
 PT to group A streptococcus and purifying the binding protein
 XX
 PS Claim 13; Page 65; 67pp; English.
 XX
 CC The patent discloses a new family of proteins termed GRAB (protein G
 CC related alpha2M binding protein) from Streptococcus pyogenes which have
 CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
 CC protein G of group G Streptococcus. GRAB protein and peptides derived
 CC from it are used in vaccine compositions for generating a protective
 CC immune response against group A Streptococcus. Antibodies against GRAB
 CC are useful for treating Streptococcus pyogenes infections. The protein
 CC is also useful for purifying alpha2M from a sample. The present sequence
 CC is a DNA encoding partial GRAB protein from S. pyogenes strain KTL3.
 CC The protein has alpha2M binding region and is useful in vaccine
 CC composition.
 CC
 XX
 SQ Sequence 504 BP; 188 A; 97 C; 108 G; 111 T; 0 other;
 Query Match 75.2%; Score 15.8; DB 21; Length 504;
 Best Local Similarity 89.5%; Pred. No. 74;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 3 actggagcacacagatctc 21
 | ||||| ||||| |||||
 Db 12 agtgggttcacacagatctc 30
 RESULT 6
 AAX30352
 ID AAX30352 standard; DNA; 697 BP.
 XX
 AC AAX30352;
 XX
 DT 14-MAY-1999 (first entry)
 XX
 DE DNA encoding a human secreted protein.
 XX
 KW Secreted protein; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; foetal deficiency; blood disorder;
 KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;
 KW renal disease; diabetes; inflammation; allergy; ischemic shock;
 KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
 KW prostate disease; asthma; osteoporosis; arthritis; ss.
 OS
 XX Homo sapiens.
 XX
 FN WO9907881-A1.
 XX
 PD 18-FEB-1999.
 PD
 XX
 PE 04-AUG-1998; 98MO-US16235.
 XX
 PR 19-AUG-1997; 97US-0056732.
 PR 05-AUG-1997; 97US-0054798.
 PR 05-AUG-1997; 97US-0054803.
 PR 05-AUG-1997; 97US-0054804.
 PR 05-AUG-1997; 97US-0054806.
 PR 05-AUG-1997; 97US-0054807.
 PR 05-AUG-1997; 97US-0054808.
 PR 05-AUG-1997; 97US-0054809.
 PR 05-AUG-1997; 97US-0055309.
 PR 05-AUG-1997; 97US-0055310.
 PR 05-AUG-1997; 97US-0055312.
 PR 05-AUG-1997; 97US-0055386.
 PR 05-AUG-1997; 97US-0055311.
 PR 18-AUG-1997; 97US-0055970.
 PR 18-AUG-1997; 97US-0055986.
 PR 18-AUG-1997; 97US-0055986.

PR 19-AUG-1997; 97US-0056365.
 PR 19-AUG-1997; 97US-0056366.
 PR 19-AUG-1997; 97US-0056557.
 PR 19-AUG-1997; 97US-0056370.
 PR 19-AUG-1997; 97US-0056371.
 PR 19-AUG-1997; 97US-0056563.
 PR 19-AUG-1997; 97US-0056731.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Ebner R, Ferlie AM, Greene JM, Janat F, Ni J;
 PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;
 XX
 DR NPI, 1999-167452/14.
 DR P-SDB; AAV10831.
 XX
 PT New isolated human genes encoding secreted polypeptides - useful for
 PT diagnosis and treatment of pathological diseases
 XX
 PS Claim 3; Page 247; 331pp; English.
 XX
 CC The specification describes secreted proteins and their corresponding
 CC polynucleotides which are useful for preventing, treating or ameliorating
 CC medical conditions, e.g. by protein or gene therapy. Pathological
 CC conditions can also be diagnosed by determining the amount of the
 CC secreted polypeptides in a sample or by determining the presence of
 CC mutations in the polynucleotides. Specific uses are described for each
 CC of the products, based on which tissues they are most highly
 CC expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental
 CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,
 CC diseases of the immune system, autoimmune diseases, hepatic and renal
 CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
 CC and cognitive disorders, schizophrenia, cardiovascular disorders,
 CC prostate diseases, asthma, disorders involving osteoclasts such as
 CC osteoporosis, arthritis or malignancies, diseases of testes, lung or
 CC thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC
 XX
 SQ Sequence 697 BP; 233 A; 121 C; 171 G; 169 T; 3 other;
 Query Match 75.2%; Score 15.8; DB 20; Length 697;
 Best Local Similarity 89.5%; Pred. No. 77;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 cactggagcacacagatc 20
 ||| ||||| ||||| |||||
 Db 274 cactggagcacacagatc 292
 RESULT 7
 AAV30310
 ID AAV30310 standard; DNA; 1073 BP.
 XX
 AC AAV30310;
 XX
 DT 28-SEP-1998 (first entry)
 XX
 DE Bacillus thuringiensis MIS-6 insecticidal toxin 196F3 DNA.
 XX
 KW Insecticide; pesticide; toxin; MIS-6; delta-endotoxin;
 KW biological control; lepidopteran; coleopteran; ss.
 OS
 XX Bacillus thuringiensis strain PS196F3 (NRRL B-21872).
 XX
 FH Key Location/Qualifiers
 FT CDS 2..1072
 FT /*tag- a
 XX
 PN WO9818932-A2.
 XX
 PD 07-MAY-1998.

XX 30-OCT-1997; 97WO-US19804.
 PF 30-OCT-1996; 96US-0029848.
 PR (MICO) MYCOGEN CORP.
 XX
 XX Dullum CJ, Fetteison JS, Loewer D, Muller-Cohn J;
 PI Narva KE, Schmeltz JL, Schnepf HE, Schwab G, Stamp L;
 PI Stockhoff BA;
 XX
 DR WPI: 1998-272226/24.
 DR P-PSDB; AAW60227.
 XX
 PT Bacillus thuringiensis isolates - used for producing pesticidal
 PT toxins and nucleotide sequences for control of lepidopterans and
 PT coleopterans
 XX
 PS Claim 5; Page 90; 139pp; English.
 XX
 CC This DNA sequence encodes a novel soluble toxin (see AAW60227) of
 CC Bacillus thuringiensis (B.t.) strain PSI96F3 (NRRL B-21872). This
 CC toxin belongs to the novel MIS-6 family of B.t. toxins that have
 CC toxicity against non-mammalian pests. The novel DNA was obtained
 CC by PCR amplification (see AAW30298-99) of total cellular genomic
 CC DNA. It can be used to produce recombinant hosts (preferably plant
 CC or bacterial) that express the toxin, or as a PCR primer or
 CC hybridization probe for use in identifying and characterizing
 CC MIS-6 family toxin genes. Disclosed and claimed are novel B.t.
 CC isolates and toxins (see AAW60218-32) which have activity against
 CC lepidopteran and/or coleopteran pests, isolated genes, probes
 CC and primers (see AAW30288-321 and AAW9734-87), and transformed host
 CC cells. The invention provides 8 entirely new families of toxins,
 CC including MIS-6, from B.t. isolates. MIS toxins are also useful
 CC for their ability to form pores in cell membranes, and can be used
 CC to facilitate entry of a second agent into a target cell.
 CC
 SO Sequence 1073 BP; 380 A; 201 C; 215 G; 270 T; 7 other;

Query Match 75.2%; Score 15.8; DB 19; Length 1073;
 Best Local Similarity 89.5%; Pred. No. 82;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 actggagatcaacagatct 21
 ||||||| |||||||
 Db 838 actggagatcaacagatct 856

RESULT 8
 AAV70397
 ID AAV70397 standard; cDNA; 4351 BP.
 XX
 AC AAV70397;
 XX
 DT 10-FEB-1999 (first entry)
 XX
 DE LRP5 isoform 2 longest open reading frame (also isoform 4,5,6).
 XX
 KM LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
 KM endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
 KM glomerulonephritis; inflammation; viral infection; osteoporosis;
 KM hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO9846743-A1.
 XX
 PD 22-OCT-1998.
 XX
 PF 15-APR-1998; 98WO-GB01102.
 XX
 PR 05-JUN-1997; 97US-0048740.
 XX

PR 15-APR-1997; 97US-0043553.
 XX
 PA (MERI) MERCK & CO INC.
 PA (WEIL) WELLCOME TRUST LTD.
 XX
 PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JM;
 PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;
 PI Phillips MS, Todd JA, Twells RCJ;
 XX
 DR WPI: 1998-594573/50.
 XX
 PT New isolated LDL-receptor related protein - used to develop products
 PT for treating e.g. elevated triglyceride levels, diabetes,
 PT autoimmune disorders, inflammation or Alzheimer's disease
 XX
 PS Claim 7; Fig 11b; 200pp; English.
 XX
 CC The present sequence represents LRP5 (low density lipoprotein (LDL)
 CC receptor related protein, previously designated LRP-3) isoform 2 cDNA.
 CC Nucleic acid molecules (NAMS) encoding LRP5 can be used for determining
 CC if an individual is susceptible to insulin dependent diabetes mellitus
 CC (IDDM). The NAMS or proteins can be used for reducing triglyceride levels
 CC in the serum of an individual. Therapies that affect LRP5 may also be
 CC useful in the treatment of autoimmune diseases such as
 CC glomerulonephritis, diseases and disorders involving disruption of
 CC endocytosis and/or antigen presentation, cytokine clearance and/or
 CC inflammation, viral infection, pathogenic bacterial toxin contamination,
 CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
 CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
 CC from the present invention can also be used for detection, diagnosis and
 CC drug screening.
 CC
 SO Sequence 4351 BP; 875 A; 1435 C; 1309 G; 732 T; 0 other;

Query Match 75.2%; Score 15.8; DB 19; Length 4351;
 Best Local Similarity 89.5%; Pred. No. 1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 cactgggatacaagatc 20
 ||||||| |||||||
 Db 1716 cactgggatacaagatc 1734

RESULT 9
 AAV70395
 ID AAV70395 standard; cDNA; 4843 BP.
 XX
 AC AAV70395;
 XX
 DT 10-FEB-1999 (first entry)
 XX
 DE LRP5 cDNA longest open reading frame.
 XX
 KM LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
 KM endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
 KM glomerulonephritis; inflammation; viral infection; osteoporosis;
 KM hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO9846743-A1.
 XX
 PD 22-OCT-1998.
 XX
 PF 15-APR-1998; 98WO-GB01102.
 XX
 PR 05-JUN-1997; 97US-0048740.
 XX
 PR 15-APR-1997; 97US-0043553.
 XX
 PA (MERI) MERCK & CO INC.
 PA (WEIL) WELLCOME TRUST LTD.
 XX

PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JH;
 PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;
 PI Phillips MS, Todd JA, Twells RCJ;
 XX
 DR WPI: 1998-594573/50.
 XX
 PT New isolated LDL-receptor related protein - used to develop products
 PT for treating, e.g. elevated triglyceride levels, diabetes,
 PT autoimmune disorders, inflammation or Alzheimer's disease
 XX
 PS Example 1, Fig 5b: 200pp; English.
 XX
 CC The present sequence represents the longest open reading frame of LRP5
 CC (low density lipoprotein (LDL) receptor related protein, previously
 CC designated LRP-3) CDNA from the present invention. Nucleic acid
 CC molecules (NMs) encoding LRP5 can be used for determining if an
 CC individual is susceptible to insulin dependent diabetes mellitus (IDDM).
 CC The NMs or proteins can be used for reducing triglyceride levels
 CC in the serum of an individual. Therapies that affect LRP5 may also be
 CC useful in the treatment of autoimmune diseases such as
 CC glomerulonephritis, diseases and disorders involving disruption of
 CC endocytosis and/or antigen presentation, cytokine clearance and/or
 CC inflammation, viral infection, pathogenic bacterial toxin contamination,
 CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
 CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
 CC from the present invention can also be used for detection, diagnosis and
 CC drug screening.
 XX
 SQ Sequence 4843 BP; 953 A; 1601 C; 1478 G; 811 T; 0 other;
 XX
 Query Match 75.2%; Score 15.8; DB 19; Length 4843;
 Best Local Similarity 89.5%; Pred. No. 1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 cactgggacacagatc 20
 Db 2207 cactgggacacagatc 2225
 XX
 RESULT 10
 ID AAV70398 standard; CDNA; 4915 BP.
 XX
 AC AAV70398;
 XX
 DT 10-FEB-1999 (first entry)
 XX
 DE LRP5 isoform 3 putative open reading frame.
 XX
 KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
 KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
 KW glomerulonephritis; inflammation; viral infection; osteoporosis;
 KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9846743-A1.
 XX
 PD 22-OCT-1998.
 XX
 PF 15-APR-1998; 98WO-GB01102.
 XX
 PR 05-JUN-1997; 97US-0048740.
 PR 15-APR-1997; 97US-0043553.
 XX
 PA (MERI) MERCK & CO INC.
 PA (WELL) WELLCOME TRUST LTD.
 XX
 PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JH;
 PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;
 PI Phillips MS, Todd JA, Twells RCJ;
 XX

DR WPI: 1998-594573/50.
 XX
 PT New isolated LDL-receptor related protein - used to develop products
 PT for treating, e.g. elevated triglyceride levels, diabetes,
 PT autoimmune disorders, inflammation or Alzheimer's disease
 XX
 PS Example 1, Fig 12c: 200pp; English.
 XX
 CC The present invention describes LRP5 (low density lipoprotein (LDL)
 CC receptor related protein, previously designated LRP-3). The present
 CC sequence represents the putative open reading frame of LRP5 isoform 3.
 CC Nucleic acid molecules (NMs) encoding LRP5 can be used for determining
 CC if an individual is susceptible to insulin dependent diabetes mellitus
 CC (IDDM). The NMs or proteins can be used for reducing triglyceride levels
 CC in the serum of an individual. Therapies that affect LRP5 may also be
 CC useful in the treatment of autoimmune diseases such as
 CC glomerulonephritis, diseases and disorders involving disruption of
 CC endocytosis and/or antigen presentation, cytokine clearance and/or
 CC inflammation, viral infection, pathogenic bacterial toxin contamination,
 CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
 CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
 CC from the present invention can also be used for detection, diagnosis and
 CC drug screening.
 XX
 SQ Sequence 4915 BP; 992 A; 1614 C; 1474 G; 835 T; 0 other;
 XX
 Query Match 75.2%; Score 15.8; DB 19; Length 4915;
 Best Local Similarity 89.5%; Pred. No. 1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 cactgggacacagatc 20
 Db 2279 cactgggacacagatc 2297
 XX
 RESULT 11
 ID AAV85549 standard; CDNA; 5022 BP.
 XX
 AC AAV85549;
 XX
 DT 10-FEB-1999 (first entry)
 XX
 DE LRP5 isoform 5 CDNA.
 XX
 KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
 KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
 KW glomerulonephritis; inflammation; viral infection; osteoporosis;
 KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9846743-A1.
 XX
 PD 22-OCT-1998.
 XX
 PF 15-APR-1998; 98WO-GB01102.
 XX
 PR 05-JUN-1997; 97US-0048740.
 PR 15-APR-1997; 97US-0043553.
 XX
 PA (MERI) MERCK & CO INC.
 PA (WELL) WELLCOME TRUST LTD.
 XX
 PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JH;
 PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;
 PI Phillips MS, Todd JA, Twells RCJ;
 XX
 DR WPI: 1998-594573/50.
 XX
 PT New isolated LDL-receptor related protein - used to develop products
 PT for treating, e.g. elevated triglyceride levels, diabetes,
 XX

PT autoimmune disorders, inflammation or Alzheimer's disease
XX
PS Claim 7; Fig 14; 200pp; English.

CC The present invention describes LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3). The present
CC sequence represents the LRP5 isoform 5 cDNA.
CC Nucleic acid molecules (NMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.

SO Sequence 5022 BP; 1036 A; 1606 C; 1503 G; 877 T; 0 other;

Query Match 75.2%; Score 15.8; DB 19; Length 5022;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 cactgggatacaagatc 20
||||| ||||| |||
DB 2178 cactgggatacaagatc 2196

RESULT 12

AAV70396
ID AAV70396 standard; cDNA; 5098 BP.

XX AAV70396;

XX 10-FEB-1999 (first entry)

DE LRP5 isoform 1 cDNA.

XX LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
XX endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
XX glomerulonephritis; inflammation; viral infection; osteoporosis;
XX hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.

OS Homo sapiens.

PN MO9846743-A1.

XX 22-OCT-1998.

PF 15-APR-1998; 98WO-GB01102.

XX 05-JUN-1997; 97US-0048740.

PR 15-APR-1997; 97US-0043553.

XX (MERI) MERCK & CO INC.

PA (WELL) WELLCOME TRUST LTD.

XX Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW;

XX Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;

XX Phillips MS, Todd JA, Twells RCJ;

XX WPI; 1998-594573/50.

PS Claim 3; Fig 5a; 200pp; English.

XX New isolated LDL-receptor related protein - used to develop products
XX for treating, e.g. elevated triglyceride levels, diabetes,
XX autoimmune disorders, inflammation or Alzheimer's disease

CC The present sequence represents LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3) isoform 1 cDNA.
CC Nucleic acid molecules (NMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.

SO Sequence 5098 BP; 1035 A; 1645 C; 1541 G; 877 T; 0 other;

Query Match 75.2%; Score 15.8; DB 19; Length 5098;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 cactgggatacaagatc 20
||||| ||||| |||
DB 2255 cactgggatacaagatc 2273

RESULT 13

AAV86019
ID AAV86019 standard; cDNA; 5125 BP.

XX AAV86019;

XX 10-FEB-1999 (first entry)

DE Lrp5 isoform 3 cDNA.

XX LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
XX endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
XX glomerulonephritis; inflammation; viral infection; osteoporosis;
XX hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.

OS Homo sapiens.

PN MO9846743-A1.

XX 22-OCT-1998.

PF 15-APR-1998; 98WO-GB01102.

XX 05-JUN-1997; 97US-0048740.

PR 15-APR-1997; 97US-0043553.

XX (MERI) MERCK & CO INC.

PA (WELL) WELLCOME TRUST LTD.

XX Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW;

XX Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;

XX Phillips MS, Todd JA, Twells RCJ;

XX WPI; 1998-594573/50.

PS Example 1; Fig 12a; 200pp; English.

XX The present invention describes LRP5 (low density lipoprotein (LDL)
XX receptor related protein, previously designated LRP-3). The present
XX sequence represents the LRP5 isoform 3 cDNA.

XX Nucleic acid molecules (NMs) encoding LRP5 can be used for determining
XX if an individual is susceptible to insulin dependent diabetes mellitus

CC (IDDM). The NAs or proteins can be used for reducing triglyceride levels
 CC in the serum of an individual. Therapies that affect LRP5 may also be
 CC useful in the treatment of autoimmune diseases such as
 CC glomerulonephritis, diseases and disorders involving disruption of
 CC endocytosis and/or antigen presentation, cytokine clearance and/or
 CC inflammation, viral infection, pathogenic bacterial toxin contamination,
 CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
 CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
 CC from the present invention can also be used for detection, diagnosis and
 CC drug screening.
 CC
 SQ Sequence 5125 BP; 1069 A; 1640 C; 1518 G; 898 T; 0 other;

Query Match 75.2%; Score 15.8; DB 19; Length 5125;
 Best Local Similarity 89.5%; Pred. No. 1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 cactgggacacagatc 20
 |||||
 DB 2282 cactgggacacagatc 2300

RESULT 14

ID AAV85550 standard; CDNA; 5162 BP.

AC AAV85550;

DT 10-FEB-1999 (first entry)

DE LRP5 isoform 6 CDNA.

XX

LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;

KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;

KM glomerulonephritis; inflammation; viral infection; osteoporosis;

KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.

XX

OS Homo sapiens.

XX

PN W09846743-A1.

XX

PD 22-OCT-1998.

XX

PF 15-APR-1998; 98WO-CB01102.

XX

PR 05-JUN-1997; 97US-0048740.

XX

PR 15-APR-1997; 97US-0043553.

XX

PA (MERI) MERCK & CO INC.

XX

PA (WELL) WELLCOME TRUST LTD.

XX

PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW;

XX

PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;

XX

PI Phillips MS, Todd JA, Twells RCU;

XX

DR WPI; 1998-594573/50.

XX

PS New isolated LDL-receptor related protein - used to develop products

XX

PT for treating, e.g. elevated triglyceride levels, diabetes,

XX

PT autoimmune disorders, inflammation or Alzheimer's disease

XX

PS Claim 7; Fig 15a; 200pp; English.

XX

PS The present invention describes LRP5 (low density lipoprotein (LDL)

XX

CC receptor related protein, previously designated LRP-3). The present

XX

CC sequence represents the LRP5 isoform 6 CDNA.

XX

CC Nucleic acid molecules (NAs) encoding LRP5 can be used for determining

XX

CC if an individual is susceptible to insulin dependent diabetes mellitus

XX

CC (IDDM). The NAs or proteins can be used for reducing triglyceride levels

XX

CC in the serum of an individual. Therapies that affect LRP5 may also be

XX

CC useful in the treatment of autoimmune diseases such as

XX

CC glomerulonephritis, diseases and disorders involving disruption of

CC endocytosis and/or antigen presentation, cytokine clearance and/or
 CC inflammation, viral infection, pathogenic bacterial toxin contamination,
 CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
 CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
 CC from the present invention can also be used for detection, diagnosis and
 CC drug screening.
 CC
 SQ Sequence 5162 BP; 1062 A; 1665 C; 1527 G; 908 T; 0 other;

Query Match 75.2%; Score 15.8; DB 19; Length 5162;
 Best Local Similarity 89.5%; Pred. No. 1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 cactgggacacagatc 20
 |||||
 DB 2319 cactgggacacagatc 2337

RESULT 15

ID AAV85551 standard; CDNA; 5166 BP.

AC AAV85551;

DT 10-FEB-1999 (first entry)

DE LRP5 isoform 2 CDNA.

XX

LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;

KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;

KM glomerulonephritis; inflammation; viral infection; osteoporosis;

KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.

XX

OS Homo sapiens.

XX

PN W09846743-A1.

XX

PD 22-OCT-1998.

XX

PF 15-APR-1998; 98WO-CB01102.

XX

PR 05-JUN-1997; 97US-0048740.

XX

PR 15-APR-1997; 97US-0043553.

XX

PA (MERI) MERCK & CO INC.

XX

PA (WELL) WELLCOME TRUST LTD.

XX

PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW;

XX

PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;

XX

PI Phillips MS, Todd JA, Twells RCU;

XX

DR WPI; 1998-594573/50.

XX

PS New isolated LDL-receptor related protein - used to develop products

XX

PT for treating, e.g. elevated triglyceride levels, diabetes,

XX

PT autoimmune disorders, inflammation or Alzheimer's disease

XX

PS Claim 8; Fig 11a; 200pp; English.

XX

PS The present invention describes LRP5 (low density lipoprotein (LDL)

XX

CC receptor related protein, previously designated LRP-3). The present

XX

CC sequence represents the LRP5 isoform 2 CDNA.

XX

CC Nucleic acid molecules (NAs) encoding LRP5 can be used for determining

XX

CC if an individual is susceptible to insulin dependent diabetes mellitus

XX

CC (IDDM). The NAs or proteins can be used for reducing triglyceride levels

XX

CC in the serum of an individual. Therapies that affect LRP5 may also be

XX

CC useful in the treatment of autoimmune diseases such as

XX

CC glomerulonephritis, diseases and disorders involving disruption of

XX

CC endocytosis and/or antigen presentation, cytokine clearance and/or

XX

CC inflammation, viral infection, pathogenic bacterial toxin contamination,

XX

CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,

XX

CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products

CC from the present invention can also be used for detection, diagnosis and
CC drug screening.

XX
SQ Sequence 5166 BP; 1082 A; 1654 C; 1530 G; 899 T; 1 other:

Query Match 75.2%; Score 15.8; DB 19; Length 5166;

Best Local Similarity 89.5%; Pred. No. 1e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cactgggacacacagatc 20

|||||

Db 2333 cactgggacacacagatc 2351

Search completed: July 25, 2001, 05:23:04
Job time: 4677 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:16:42 ; Search time 1290.33 Seconds

(without alignments)
251.736 Million cell updates/sec

Title: US-09-142-095-2

Perfect score: 21

Sequence: 1 ccactggatcacaagatctct 21

Scoring table: IDENTITY_NUC

Searched: 1344157 seqs, 7733874588 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database :

Listing first 45 summaries

1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_higo_hum:*
20: em_higo_inv:*
21: em_higo_rod:*
22: em_hig_hum1:*
23: em_hig_hum2:*
24: em_hig_hum3:*
25: em_hig_hum4:*
26: em_hig_hum5:*
27: em_hig_hum6:*
28: em_hig_hum7:*
29: em_hig_hum8:*
30: em_hig_inv1:*
31: em_hig_inv2:*
32: em_hig_other:*
33: em_hig_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_rod:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_hig1:*
61: gb_hig2:*
62: gb_hig3:*
63: gb_hig4:*
64: gb_hig5:*
65: gb_hig6:*
66: gb_hig7:*
67: gb_hig8:*
68: gb_hig9:*
69: gb_hig10:*
70: gb_hig11:*
71: gb_hig12:*
72: gb_hig13:*
73: gb_hig14:*
74: gb_hig15:*
75: gb_hig16:*
76: gb_hig17:*
77: gb_hig18:*
78: gb_hig19:*
79: gb_hig20:*
80: gb_hig21:*
81: gb_hig22:*
82: gb_hig23:*
83: gb_hig24:*
84: gb_hig25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rod:*
95: gb_rod2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	21	100.0	21	9 A65501	A65501 Sequence 2
2	21	100.0	292	89 AF357220	AF357220 Homo sapi
3	21	100.0	531	89 AF352795	AF352795 Homo sapi
4	21	100.0	541	89 AF180372	AF180372 Homo sapi
5	21	100.0	918	88 AF110194	AF110194 Homo sapi
6	21	100.0	1190	97 HOMOG71A	M84125 Human billir
7	21	100.0	2351	97 HOMOG71A	M57899 Human billir
8	21	100.0	3341	91 D87674	D87674 Homo sapien

```

RESULT 2
AF357220/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
Mokris, K., Addington, T., Nguyen, T.S., Glendening, M., Kutlar, F. and
Kutlar, A.
TITLE
Detection of TATA box TA repeat region [6(TA)repeat] of human
bllitubin UDP-glucuronosyltransferase 1-1 gene (UGT1L1) in an
African American individual
Unpublished
JOURNAL
AUTHORS
Mokris, K., Addington, T., Nguyen, T.S., Glendening, M., Kutlar, F. and
Kutlar, A.
TITLE
Direct Submission
SUBMITTED
(07-MAR-2001) Medicine/Hemoglobin DNA Laboratory, Medical
College of Georgia, 15th Street, Augusta, GA 30912, USA
FEATURES
SOURCE
1.292
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37"
/cell_type="WBC"
/tissue_type="whole blood"
1..292
/gene="UGT1"

```

RESULT 3
AF352795/c 531 bp DNA PRI 10-APR-2001
LOCUS Homo sapiens bilirubin UDP-glucuronosyltransferase 1-1 (UGT1A1) gene,
DEFINITION UGT1A1*1 allele, partial cds.
ACCESSION AF352795
VERSION AF352795.1 GI:13569708
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
TITLE 7(f)a repeat polymorphism of the TATA box of human bilirubin
UDP-glucuronosyltransferase 1-1(UGT1A1*1) gene in a patient with
sickle cell anemia + high bilirubinemia
JOURNAL Unpublished
2 (bases 1 to 531)
REFERENCE McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
AUTHORS Direct Submission
TITLE Submitted (23-FEB-2001) Medicine/Hemoglobin DNA Laboratory, Medical
JOURNAL College of Georgia, 15th St., AC-1000, Augusta, GA 30912, USA
FEATURES
source
location/Qualifiers
1..531
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37"
/cell_type="white blood cell"
/tissue_type="blood"
1..531
/gene="UGT1A1"
/note="UGT1"
/allele="UGT1A1*1"
1..313
/gene="UGT1A1"
279..295
/gene="UGT1A1"
279..292
repeat_region
/note="repeat polymorphism compared to UGT1A1 sequence
presented in Genbank Accession Number AF180372; contains 7
ta repeats frequently detected in patients with high
bilirubinemia"
/rpt_type=tandem
/rpt_unit=ta
314..531
/gene="UGT1A1"
/product="bilirubin UDP-glucuronosyltransferase 1-1"
334..531
/gene="UGT1A1"
/codon_start=1
/product="bilirubin UDP-glucuronosyltransferase 1-1"
/protein_id="AAK31204.1"
/db_xref="GI:13569708"
/translation="MAVESOGGRPLVGLLCLVLPVVSAGKILLIPVDSHWSML
GAIOOLQORGEHYVLAPDASLYIRG"
BASE COUNT 120 a 121 c 137 g 153 t
ORIGIN

Query Match 100.0%; Score 21; DB 89; Length 531;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccactggatcaacagctatct 21
|||||
DB 439 CCACGGGATCAACAGTACT 419

DEFINITION Homo sapiens bilirubin UDP-glucuronosyltransferase 1-1 (UGT1A1) gene,
UGT1A1*1 allele, partial cds.
ACCESSION AF180372
VERSION AF180372.1 GI:6010649
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Kutlar,F., Stomek,E., Leitner,C., Nechtman,J. and Kutlar,A.
TITLE Detection of the TATA box polymorphism of the human bilirubin
UDP-glucuronosyltransferase 1-1 gene (UGT1A1) in a patient with
sickle cell anemia
JOURNAL Unpublished
2 (bases 1 to 541)
REFERENCE Kutlar,F., Stomek,E., Leitner,C., Nechtman,J. and Kutlar,A.
AUTHORS Direct Submission
TITLE Submitted (24-AUG-1999) Medicine, Hematology/Oncology-Sickle Cell
JOURNAL Center, Medical College of Georgia, 15th Street, AC-1000, Augusta,
GA 30912, USA
FEATURES
source
location/Qualifiers
1..541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37"
/cell_type="white blood cells"
/tissue_type="whole blood"
1..541
/gene="UGT1"
/note="GNT1; UGT1A1"
/allele="UGT1A1*1"
1..313
/gene="UGT1"
226
/gene="UGT1"
/replace="t"
279..288
/note="polymorphic region"
/rpt_type=tandem
/rpt_unit=ta
279..291
/gene="UGT1"
/note="Ritter,J.R., et al., 1992, J. Biol. Chem.,
267:3257-3261"
314..541
/gene="UGT1"
/product="bilirubin UDP-glucuronosyltransferase 1-1"
330..541
/gene="UGT1"
/note="UDP-glucosyltransferase 1"
/codon_start=1
/product="bilirubin UDP-glucuronosyltransferase 1-1"
/protein_id="AAK31205.1"
/db_xref="GI:6010650"
/translation="MAVESOGGRPLVGLLCLVLPVVSAGKILLIPVDSHWSML
GAIOOLQORGEHYVLAPDASLYIRG"
BASE COUNT 123 a 124 c 141 g 153 t
ORIGIN

Query Match 100.0%; Score 21; DB 89; Length 541;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccactggatcaacagctatct 21
|||||
DB 435 CCACGGGATCAACAGTACT 415

RESULT 4
AF180372/c 541 bp DNA PRI
LOCUS AF180372 541 bp DNA PRI
DEFINITION AF180372 541 bp DNA PRI
ACCESSION AF180372
VERSION AF180372.1
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Kutlar,F., Stomek,E., Leitner,C., Nechtman,J. and Kutlar,A.
TITLE Detection of the TATA box polymorphism of the human bilirubin
UDP-glucuronosyltransferase 1-1 gene (UGT1A1) in a patient with
sickle cell anemia
JOURNAL Unpublished
2 (bases 1 to 541)
REFERENCE Kutlar,F., Stomek,E., Leitner,C., Nechtman,J. and Kutlar,A.
AUTHORS Direct Submission
TITLE Submitted (24-AUG-1999) Medicine, Hematology/Oncology-Sickle Cell
JOURNAL Center, Medical College of Georgia, 15th Street, AC-1000, Augusta,
GA 30912, USA
FEATURES
source
location/Qualifiers
1..541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37"
/cell_type="white blood cells"
/tissue_type="whole blood"
1..541
/gene="UGT1"
/note="GNT1; UGT1A1"
/allele="UGT1A1*1"
1..313
/gene="UGT1"
226
/gene="UGT1"
/replace="t"
279..288
/note="polymorphic region"
/rpt_type=tandem
/rpt_unit=ta
279..291
/gene="UGT1"
/note="Ritter,J.R., et al., 1992, J. Biol. Chem.,
267:3257-3261"
314..541
/gene="UGT1"
/product="bilirubin UDP-glucuronosyltransferase 1-1"
330..541
/gene="UGT1"
/note="UDP-glucosyltransferase 1"
/codon_start=1
/product="bilirubin UDP-glucuronosyltransferase 1-1"
/protein_id="AAK31205.1"
/db_xref="GI:6010650"
/translation="MAVESOGGRPLVGLLCLVLPVVSAGKILLIPVDSHWSML
GAIOOLQORGEHYVLAPDASLYIRG"
BASE COUNT 123 a 124 c 141 g 153 t
ORIGIN

RESULT 5
AF10194/c

LOCUS AF110194 918 bp DNA PRI 02-JAN-2001
 DEFINITION Homo sapiens chromosome 2 UDP-glucuronosyltransferase (UGT1A1) gene, UGT1A1*33 allele, partial cds.
 ACCESSION AF110194
 VERSION AF110194.1 GI:12002134
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 918)
 AUTHORS Guillemette, C.
 TITLE Direct Submission
 JOURNAL Submitted (01-DEC-1998) Center for Cancer Research, MIT, 77 Massachusetts Avenue, E17-540, Cambridge, MA 02139, USA
 FEATURES
 source
 1. 918
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 <38..>918
 /gene="UGT1A1"
 /product="UDP-glucuronosyltransferase"
 <38..>918
 /gene="UGT1A1"
 /note="22..918; 1A1 variant allele; L233R"
 /allele="UGT1A1*33"
 38..>918
 /gene="UGT1A1"
 /codon_start=1
 /product="UDP-glucuronosyltransferase"
 /protein_id="AA63197.1"
 /db_xref="GI:12002135"
 /translation="MAVESQGRPLVGLLCEVGPVSHAKILLIPDGSMLSL
 GAIQLOQRGEHIVLAPDASLYIRDGAFTLKTYPVPRQREDEVESFVSGHNVEN
 DSFLQRIKTKYKKIKKDSAMLSCGSHLHNKELMASLESSFDVMTLDPFLCSPIT
 AYLSTPVEFLHAPCSLEFEKTCPCPFYPRPLSSSDMTFLQVKNMLIAS
 QNFLCDVYSPTATRASEFLQREYVQDLSSASVWLFPSDVKYPRIPMPNMFVG
 GINCHONPESQVIGVG"
 BASE COUNT 196 a 245 c 235 g 242 t
 ORIGIN
 Query Match 100.0%; Score 21; DB 88; Length 918;
 Best Local Similarity 100.0%; Pred. No. 0.74;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

TITLE A novel complex locus UGT1 encodes human bilirubin, phenol, and other UDP-glucuronosyltransferase isozymes with identical carboxyl terminal
 JOURNAL J. Biol. Chem. 267 (5), 3257-3261 (1992)
 MEDLINE 92147680
 FEATURES
 source
 Location/Qualifiers
 1. 1190
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="liver"
 /tissue_lib="cosmid"
 /lab_host="XLI-Blue"
 32..46
 /gene="UGT1A"
 32..46
 /gene="UGT1A"
 69..948
 /gene="UGT1A"
 /number=1
 69..1190
 /gene="UGT1A"
 85..>948
 /gene="UGT1A"
 /codon_start=1
 /product="Bilirubin UDP-glucuronosyltransferase"
 /protein_id="AA61248.1"
 /db_xref="GI:340132"
 /translation="MAVESQGRPLVGLLCEVGPVSHAKILLIPDGSMLSL
 GAIQLOQRGEHIVLAPDASLYIRDGAFTLKTYPVPRQREDEVESFVSGHNVEN
 DSFLQRIKTKYKKIKKDSAMLSCGSHLHNKELMASLESSFDVMTLDPFLCSPIT
 AYLSTPVEFLHAPCSLEFEKTCPCPFYPRPLSSSDMTFLQVKNMLIAS
 QNFLCDVYSPTATRASEFLQREYVQDLSSASVWLFPSDVKYPRIPMPNMFVG
 GINCHONPESQVIGVG"
 949..1190
 /partial
 /gene="UGT1A"
 /note="does not fit consensus"
 /number=1
 BASE COUNT 261 a 290 c 286 g 353 t
 ORIGIN
 Query Match 100.0%; Score 21; DB 97; Length 1190;
 Best Local Similarity 100.0%; Pred. No. 0.74;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

/db_xref="taxon:9606"
/dev_stage="adult"
/sex="female"
/tissue_type="liver"
/tissue_id="lambda-zap"
/map="2"
16..1617
/gene="UGT1"
16..1617
/gene="UGT1"
/standard_name="bilirubin UDP-glucuronosyltransferase
isozyme 1"
/EC_number="2.4.1.17"
/codon_start=1
/db_xref="GDB:600-120-007"
/product="UDP-glucuronosyltransferase 1"
/protein_id="AA63195.1"
/db_xref="GI:184473"
/translation="MAVESOGGRPLVGLLCVLPVSHAKILLIPVDSHWLSML
CAIOQDQKREHYVLPDASLYIRGAFITLTPYFQREDEKESFVSLGHVFN
DSFLQVNTIKTKIKDSAMLSGSHLHNKELMASIESFDVMTLPFLPCSPIY
AOYLSLPVPELHALPCSLFEATQCPNPFSSVPRPLSHSDHMTLPQVKMLIAFS
ONELCPVSPYATLASEFLQREVTQDLSASVWLEFSDPKVDYPRDIPNMFVFG
GINCLHONPLSQEPAYINAGSEHGIYVSGSVSEIPEKKMAIADAKITQTVL
WRYGTBPSNLANTLLVWLPONDILGHPMTAFTHAGSHGVESICNCPVPMAP
LPEDONKAKEMERKGVNTLNTERTSEDEENAKAVINDKSKENIKRSLHKDR
PVEPDLAVFVETFMHKGAPLIRPAHDLTWYQHSLOVIFLLAVLVAFITTEK
CCAYGRKCLGKRGVRAKRSKTH"
43..75
/gene="UGT1"
/note="G00-120-007"
76..1614
/gene="UGT1"
/standard_name="bilirubin UDP-glucuronosyltransferase
isozyme 1"
/EC_number="2.4.1.17"
/note="G00-120-007"
/product="UDP-glucuronosyltransferase 1"
BASE COUNT      601 a      540 c      557 g      653 t
ORIGIN

```

```

Query Match      100.0%; Score 21; DB 97; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ccactggatcacagatctc 21
|||||
Db      121 CCACGGGATCACAGATCT 101

```

```

RESULT 8
D87674/c 3341 bp DNA PRI 14-APR-2000
LOCUS Homo sapiens gene for bilirubin UDP-glucuronosyltransferase 1,
DEFINITION promoter region and partial cds.
ACCESSION D87674.1 GI:3059176
VERSION bilirubin UDP-glucuronosyltransferase 1.
KEYWORDS Homo sapiens DNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 3341)
AUTHORS Ueyama, H., Kotani, O., Soeda, Y., Sato, H., Sato, Y., Okubo, I. and
Doi, Y.
TITLE Analysis of the promoter of human bilirubin
UDP-glucuronosyltransferase gene (UGT1) in relevance to Gilbert's
syndrome
JOURNAL Hepatol. Res. 9, 152-163 (1997)
REFERENCE 2 (bases 1 to 3341)
AUTHORS Ueyama, H.

```

```

TITLE Direct Submission
JOURNAL Submitted (04-SEP-1996) to the DDBJ/EMBL/GenBank databases. Hisao
Ueyama, Shiga University of Medical Science, Department of Medical
Biochemistry; Seta, Otsu, Shiga 520-21, Japan (Tel:077-548-2162,
Fax:077-548-2164)
COMMENT Sequence updated (08-Jan-1997) by: Hisao Ueyama.
FEATURES
source
1..3341
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2q37"
misc_feature
2538..2544
/note="XRE"
misc_feature
2606..2610
/note="XRE"
misc_feature
3088..3097
/note="E-box"
misc_feature
3101..3113
/note="HNF-1 site"
CAAT_signal
3125..3129
TATA_signal
3177..3341
exon
3177..3341
number=1
gene
3192..3341
/gene="UGT1*1"
3192..3341
/gene="UGT1*1"
/codon_start=1
/product="bilirubin UDP-glucuronosyltransferase 1"
/protein_id="BA25600.1"
/db_xref="GI:3059177"
/translation="MAVESOGGRPLVGLLCVLPVSHAKILLIPVDSHWLSML
GAIQDL"
BASE COUNT      893 a      695 c      803 g      950 t
ORIGIN

```

```

Query Match      100.0%; Score 21; DB 91; Length 3341;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ccactggatcacagatctc 21
|||||
Db      3297 CCACGGGATCACAGATCT 3277

```

```

RESULT 9
AC006985 68770 bp DNA PRI 21-DEC-1999
LOCUS Homo sapiens BAC clone RP11-154L24 from 2, complete sequence.
DEFINITION AC006985
ACCESSION AC006985
VERSION AC006985.2 GI:5732165
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 68770)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
2 (bases 1 to 68770)
REFERENCE Gattung, S., Stoneking, T. and Davidson, T.
TITLE The sequence of Homo sapiens BAC clone RP11-154L24
JOURNAL unpublished
AUTHORS 3 (bases 1 to 68770)
REFERENCE Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 68770)

```

AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (13-AUG-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 68770)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 68770)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Aug 13, 1999 this sequence version replaced gi:4337256.

Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: saplens@wustl.wustl.edu
 Summary Statistics
 Center project name: H_NH0154124

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frenken, E., Tatenio, M., Catanesse, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://daccpac.med.buffalo.edu>)
 VECTOR: pBACpac.med.buffalo.edu

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-332L11, 200 bp overlap. Actual start of this clone is at base position 8614 of RP11-332L11; actual end is at base position 68770 of RP11-154L24.

The clone RP11-154L24 contains a tandem repeat from base positions 38234 to 39039, this region contains some low quality data. The assembly is consistent with the restriction digest information.

FEATURES

SOURCE
 1. 68770
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone="RP11-154L24"
 3. 192
 /clone_lib="RPCT-11"
 /rpt_family="L1"
 repeat_region
 221. 528

repeat_region
 529. 1634
 /rpt_family="L1"
 repeat_region
 1638. 1878
 /rpt_family="MIR"
 repeat_region
 2079. 2271
 /rpt_family="MERL_type"
 repeat_region
 2276. 2336
 /rpt_family="MERL_type"
 repeat_region
 4683. 4806
 /rpt_family="MIR"
 repeat_region
 5131. 5389
 /rpt_family="MALR"
 repeat_region
 8738. 8871
 /rpt_family="MIR"
 repeat_region
 9092. 9371
 /rpt_family="AchoDo"
 repeat_region
 9967. 10582
 /rpt_family="L1"
 repeat_region
 10576. 11200
 /rpt_family="L1"
 repeat_region
 11196. 11469
 /rpt_family="L1"
 repeat_region
 11540. 11654
 /rpt_family="L1"
 repeat_region
 12108. 12523
 /rpt_family="MALR"
 repeat_region
 13113. 13149
 /rpt_family="MERL_type"
 repeat_region
 13150. 13171
 /rpt_family="(CAG)n"
 repeat_region
 13172. 13286
 /rpt_family="MERL_type"
 repeat_region
 13338. 13497
 /rpt_family="MERL_type"
 repeat_region
 14727. 14848
 /rpt_family="MIR"
 repeat_region
 15890. 16244
 /rpt_family="MALR"
 repeat_region
 16380. 16723
 /rpt_family="L1"
 repeat_region
 17483. 17560
 /rpt_family="L2"
 repeat_region
 17655. 17747
 /rpt_family="L2"
 repeat_region
 18220. 18337
 /rpt_family="MERL_type"
 repeat_region
 18770. 18816
 /rpt_family="(CA)n"
 repeat_region
 19103. 19121
 /rpt_family="AT-rich"
 repeat_region
 19122. 19636
 /rpt_family="L1"
 repeat_region
 19637. 19662
 /rpt_family="AT-rich"
 repeat_region
 19825. 20014
 /rpt_family="L1"
 repeat_region
 20034. 20098
 /rpt_family="MIR"
 repeat_region
 20254. 21067
 /rpt_family="L1"
 repeat_region
 21441. 21162
 /rpt_family="A-rich"
 repeat_region
 24649. 25018
 /rpt_family="Retroviral"
 repeat_region
 25258. 25447
 /rpt_family="MIR"
 repeat_region
 25777. 25824
 /rpt_family="(CA)n"
 repeat_region
 28020. 28178
 /rpt_family="MIR"
 repeat_region
 28178. 28232
 /rpt_family="L2"

```

repeat_region 28503..28558 /rpl_family="L1"
repeat_region 28558..28741 /rpl_family="L1"
repeat_region 28741..29064 /rpl_family="L1"
repeat_region 29064..29276 /rpl_family="L1"
repeat_region 29276..29434 /rpl_family="MERL-type"
repeat_region 29434..29712 /rpl_family="MERL-type"
repeat_region 29712..30867 /rpl_family="MERL-type"
repeat_region 30867..31015 /rpl_family="MERL-type"
repeat_region 31015..31268 /rpl_family="MERL-type"
repeat_region 31268..31662 /rpl_family="MERL-type"
repeat_region 31662..32195 /rpl_family="MALR"
repeat_region 32195..32390 /rpl_family="MALR"
repeat_region 32390..33607 /rpl_family="MALR"
repeat_region 33607..33760 /rpl_family="L2"
repeat_region 33760..34127 /rpl_family="L2"
repeat_region 34127..34167 /rpl_family="Alu"
repeat_region 34167..34283 /rpl_family="GA-rich"
repeat_region 34283..34806 /rpl_family="GA-rich"
repeat_region 34806..34976 /rpl_family="L2"
repeat_region 34976..35247 /rpl_family="L2"
repeat_region 35247..35719 /rpl_family="L2"
repeat_region 35719..35762 /rpl_family="L2"
repeat_region 35762..35763 /rpl_family="GA-rich"
repeat_region 35763..35909 /rpl_family="GA-rich"

```

```

Query Match 100.0% Score 21: DB 86; Length 68770;
Best Local Similarity 100.0% Pred. No. 0.81:
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0:
QY 1 cccacggagatcaacagatct 21
Db 66050 CCNCTGGATCAACAGATCT 66070

```

```

RESULT 10
AC026497 176619 bp DNA HTG 22-MAR-2000
LOCUS Homo sapiens chromosome 11 clone RP11-689A10 map 11, *** SEQUENCING
DEFINITION AC026497.1 GI:7283186
IN PROGRESS ***; 43 unordered pieces.
ACCESSION AC026497
VERSION AC026497.1
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 176619)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abrahams,H., Allen,N.,
Anderson,S., Baldwin,J., Batra,N., Bastien,V., Beda,F.,
Boguski,M.S., Bouckhelter,B., Brown,A., Burtett,G.,
Campopiano,A., Castle,A., Choepey,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domingo,M., Doyle,M., Fereira,P., Fitzhugh,W., Gage,D.,
Galegher,J., Gardner,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heath,A., Horton,L.,
Holland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karas,A.,
Klein,J., Lacombe,K., Lamazares,R., Landers,T., Lebecky,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N.,
McCarthy,M., McGowan,P., McGuire,A., McKernan,K., McPherson,R.,
Meldrum,J., Menees,L., Mihov,T., Miranda,C., Minger,A., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisanti,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

```

TITLE JOURNAL COMMENT

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,K.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
<http://ftp.genome.washington.edu/RN/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 16277
Center clone name: 689A10

NOTE: This is a 'working draft' sequence. It currently
consists of 43 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

```

1 1333: contig of 1333 bp in length
1334 1433: gap of 100 bp
1434 2787: contig of 1354 bp in length
2788 2887: gap of 100 bp
2888 3998: contig of 1111 bp in length
3999 4098: gap of 100 bp
4099 5777: contig of 1679 bp in length
5778 5877: gap of 100 bp
5878 7685: contig of 1808 bp in length
7686 7785: gap of 100 bp
7786 8176: contig of 391 bp in length
8177 8276: gap of 100 bp
8277 9926: contig of 1650 bp in length
9927 10026: gap of 100 bp
10027 11498: contig of 1473 bp in length
11499 11599: gap of 100 bp
11600 13616: contig of 2017 bp in length
13617 13716: gap of 100 bp
13717 16073: contig of 2357 bp in length
16074 16173: gap of 100 bp
16174 18510: contig of 2337 bp in length
18511 18610: gap of 100 bp
18611 20495: contig of 1885 bp in length
20496 20595: gap of 100 bp
20596 23548: contig of 2953 bp in length
23549 23648: gap of 100 bp
23649 25910: contig of 2262 bp in length
25911 26010: gap of 100 bp
26011 28060: contig of 2050 bp in length
28061 28160: gap of 100 bp
28161 30134: contig of 1974 bp in length
30135 30234: gap of 100 bp
30235 32646: contig of 2412 bp in length
32647 32746: gap of 100 bp
32747 33096: contig of 2350 bp in length
33097 35196: gap of 100 bp
35197 37966: contig of 2500 bp in length
37967 37996: gap of 100 bp
37997 40539: contig of 2743 bp in length
40540 40639: gap of 100 bp
40640 45518: contig of 4879 bp in length
45519 45618: gap of 100 bp
45619 48728: contig of 3110 bp in length
48729 48828: gap of 100 bp
48829 53099: contig of 3271 bp in length
53100 53199: gap of 100 bp

```



```

* 52200 56747: contig of 4548 bp in length
* 56748 56647: gap of 100 bp
* 56648 60467: contig of 3620 bp in length
* 60468 60567: gap of 100 bp
* 60568 64947: contig of 4380 bp in length
* 64948 65047: gap of 100 bp
* 65048 69431: contig of 4384 bp in length
* 69432 69531: gap of 100 bp
* 69532 74167: contig of 4636 bp in length
* 74168 74267: gap of 100 bp
* 74268 78188: contig of 3921 bp in length
* 78189 78288: gap of 100 bp
* 78289 81664: contig of 3376 bp in length
* 81665 81764: gap of 100 bp
* 81765 85169: contig of 3405 bp in length
* 85170 85269: gap of 100 bp
* 85270 90445: contig of 5176 bp in length
* 90446 90545: gap of 100 bp
* 90546 95008: contig of 4463 bp in length
* 95009 95108: gap of 100 bp
* 95109 100587: contig of 5479 bp in length
* 100588 100687: gap of 100 bp
* 100688 105884: contig of 5197 bp in length
* 105885 105984: gap of 100 bp
* 105985 111486: contig of 5502 bp in length
* 111487 111586: gap of 100 bp
* 111587 117781: contig of 6195 bp in length
* 117782 117881: gap of 100 bp
* 117882 125343: contig of 7462 bp in length
* 125344 125443: gap of 100 bp
* 125444 132508: contig of 7065 bp in length
* 132509 132608: gap of 100 bp
* 132609 141190: contig of 8582 bp in length
* 141191 141290: gap of 100 bp
* 141291 152556: contig of 11266 bp in length
* 152557 152656: gap of 100 bp
* 152657 163590: contig of 10934 bp in length
* 163591 163690: gap of 100 bp
* 163691 176619: contig of 12929 bp in length.
*
FEATURES
Source
1..176619
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-689A10"
/clone_id="RPC1-11 Human Male BAC"
1..1333
/note="assembly-fragment"
misc_feature
1434..2787
/note="assembly-fragment"
misc_feature
2888..3998
/note="assembly-fragment"
misc_feature
4099..5777
/note="assembly-fragment"
misc_feature
5878..7685
/note="assembly-fragment"
misc_feature
7786..8176
/note="assembly-fragment"
misc_feature
8277..9926
/note="assembly-fragment"
misc_feature
10027..11493
/note="assembly-fragment"
misc_feature
11600..11616
/note="assembly-fragment"
misc_feature
13717..16073
/note="assembly-fragment"
misc_feature
16174..18510
/note="assembly-fragment"
misc_feature
18611..20495
/note="assembly-fragment"

```

```

misc_feature 20596..22548
/note="assembly-fragment"
misc_feature 23649..25910
/note="assembly-fragment"
misc_feature 26011..28060
/note="assembly-fragment"
misc_feature 28161..30134
/note="assembly-fragment"
misc_feature 30235..32646
/note="assembly-fragment"
misc_feature 32747..35096
/note="assembly-fragment"
misc_feature 35197..37696
/note="assembly-fragment"
misc_feature 37797..40539
/note="assembly-fragment"
misc_feature 40640..45518
/note="assembly-fragment"
misc_feature 45619..48728
/note="assembly-fragment"

Query Match 100.0%; Score 21; DB 70; Length 176619;
Best Local Similarity 100.0%; Pred. No. 0.83; 0; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;

QY 1 ccactggatcacagatct 21
Db 89296 CCACGTGGATCACAGTATCT 89316

RESULT 11
AF297093 198872 bp DNA PRI 07-NOV-2000
LOCUS AF297093/c
DEFINITION Homo sapiens UGT1 gene locus, complete sequence.
ACCESSION AF297093
VERSION AF297093.1 GI:11118740
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 198872)
AUTHORS Owens,I.S., Gong,Q., Cho,J.W., Huang,T., Potter,C., Gholami,N.,
Basu,N.K., Kubota,S., Carvalho,S. and Pennington,M.W.
Thirteen UDP glucuronosyltransferase genes encoded at the human
UGT1 locus
JOURNAL Unpublished
2 (bases 1 to 198872)
REFERENCE Owens,I.S., Gong,Q., Cho,J.W., Huang,T., Potter,C., Gholami,N.,
Basu,N.K., Kubota,S., Carvalho,S. and Pennington,M.W.
Direct Submission
JOURNAL Submitted (10-AUG-2000) Heritable Disorders Branch, NIH-NICHD, 9000
Rockville Pike, Bethesda, MD 20892, USA
FEATURES
Source
1..198872
/organism="Homo sapiens"
/db_xref="taxon:9606"
1989..2841
/gene="UGT1A2p"
/note="UDP glucuronosyltransferase 1A12"
pseudo
1989..2841
/gene="UGT1A2p"
19970..20956
/codon_start=1
/pseudo
19970..20956
/gene="UGT1A1p"
/note="UDP glucuronosyltransferase 1A11"
pseudo
19970..19978
TATA_signal
CDS 20105..20956

```

/gene="UGT1A11p"
 /codon_start=1
 /pseudo
 34133..34149
 /gene="UGT1A6"
 /gene="UGT1A8"
 34133..>187313
 /gene="UGT1A10"
 join(34264..35118,181788..181919,182603..182690,
 182974..183193,187016..>187313)
 /gene="UGT1A8"
 /product="UDP glucuronosyltransferase 1A8"
 join(34264..35118,181788..181919,182603..182690,
 182974..183193,187016..187313)
 /gene="UGT1A8"
 /codon_start=1
 /product="UDP glucuronosyltransferase 1A8"
 /protein_id="AAG30416.1"
 /db_xref="GI:11118741"
 /translation="MARTGWTSPILCVSLITCGFAEAGKLLVPMDSGSHMTNOSY
 VEKILRGHEVYVYVPEVSMOLGSLNCTVKTSTYTLDDREFPFAHAKMAY
 RSLFSLTSSNGFNLFFSHCRSLFNDKKEVLEKESFDVFLDPDAGLIVAKY
 FSLPSVFRAGILCHYLEGACCPAPLSVPRILLGSDAMFKERVNNHMLIEHL
 FCGYSKNALEISEILQTPYVDYLSHSISMLRTDVPDYDKRYVNMFTIGIN
 CHGKPLPMEFEATINASGEHGIYVFSLSGSMSEIPEKKAMAIDALGKIPOYVMRY
 TGRPSNLANTTILVKNLPONDILGHPTMRAITHAGSHVETISICNGVPMWPLFG
 DOMNAKMETKAGVTLVNLEMTSEDLNALKAVINDSKYENIMRLSSLHKDRPE
 PLDAVFWVEFVRHKGAPHLRPAHDLTWYQYHSLDVGFLAVLVAVTFAITFKCA
 YGRCGLGKGRVKAKHKSXT"
 52972..52988
 /gene="UGT1A10"
 52972..>187313
 /gene="UGT1A10"
 join(53103..53957,181788..181919,182603..182690,
 182974..183193,187016..>187313)
 /gene="UGT1A10"
 /product="UDP glucuronosyltransferase 1A10"
 join(53103..53957,181788..181919,182603..182690,
 182974..183193,187016..187313)
 /gene="UGT1A10"
 /codon_start=1
 /product="UDP glucuronosyltransferase 1A10"
 /protein_id="AAG30417.1"
 /db_xref="GI:11118742"
 /translation="MARAGWTSPVPLCVCLITCGFAEAGKLLVPMDSGSHMTNOSY
 VEKILRGHEVYVYVPEVSMOLGSLNCTVKTSTYTLDDREFPFAHAKMAY
 OSIFSLTSSNGFNLFFSHCRSLFNDKKEVLEKESFDVFLDPDAGLIVAKY
 FSLPSVFRAGILCHYLEGACCPAPLSVPRILLGSDAMFKERVNNHMLIEHL
 FCGYSKNALEISEILQTPYVDYLSHSISMLRTDVPDYDKRYVNMFTIGIN
 CHGKPLPMEFEATINASGEHGIYVFSLSGSMSEIPEKKAMAIDALGKIPOYVMRY
 TGRPSNLANTTILVKNLPONDILGHPTMRAITHAGSHVETISICNGVPMWPLFG
 DOMNAKMETKAGVTLVNLEMTSEDLNALKAVINDSKYENIMRLSSLHKDRPE
 PLDAVFWVEFVRHKGAPHLRPAHDLTWYQYHSLDVGFLAVLVAVTFAITFKCA
 YGRCGLGKGRVKAKHKSXT"
 64509..65608
 /gene="UGT1A13p"
 /note="UDP glucuronosyltransferase 1A13"
 /pseudo
 join(64509..65011,65353..65608)
 /gene="UGT1A13p"
 /codon_start=1
 /pseudo
 88416..>187313
 /gene="UGT1A9"
 88416..88427
 /gene="UGT1A9"
 join(88543..89397,181788..181919,182603..182690,
 182974..183193,187016..>187313)
 /gene="UGT1A9"
 /product="UDP glucuronosyltransferase 1A9"
 join(88543..89397,181788..181919,182603..182690,
 182974..183193,187016..187313)
 /gene="UGT1A9"
 /codon_start=1

/product="UDP glucuronosyltransferase 1A9"
 /protein_id="AAG30418.1"
 /db_xref="GI:11118743"
 /translation="MACGWTSPVPLCVCLITCGFAEAGKLLVPMDSGSHMTNOSY
 VEKILRGHEVYVYVPEVSMOLGSLNCTVKTSTYTLDDREFPFAHAKMAY
 RSIXSLTSSNGFNLFFSHCRSLFNDKKEVLEKESFDVFLDPDAGLIVAKY
 FSLPSVFRAGILCHYLEGACCPAPLSVPRILLGSDAMFKERVNNHMLIEHL
 LCHRRFVNALEISEILQTPYVDYLSHSISMLRTDVPDYDKRYVNMFTIGIN
 CHGKPLPMEFEATINASGEHGIYVFSLSGSMSEIPEKKAMAIDALGKIPOYVMRY
 TGRPSNLANTTILVKNLPONDILGHPTMRAITHAGSHVETISICNGVPMWPLFG
 DOMNAKMETKAGVTLVNLEMTSEDLNALKAVINDSKYENIMRLSSLHKDRPE
 PLDAVFWVEFVRHKGAPHLRPAHDLTWYQYHSLDVGFLAVLVAVTFAITFKCA
 YGRCGLGKGRVKAKHKSXT"
 98424..98438
 /gene="UGT1A7"
 98424..>187313
 /gene="UGT1A7"
 join(98552..99406,181788..181919,182603..182690,
 182974..183193,187016..>187313)
 /gene="UGT1A7"
 /product="UDP glucuronosyltransferase 1A7"
 join(98552..99406,181788..181919,182603..182690,
 182974..183193,187016..187313)
 /gene="UGT1A7"
 /codon_start=1
 /product="UDP glucuronosyltransferase 1A7"
 /protein_id="AAG30419.1"
 /db_xref="GI:11118744"
 /translation="MARAGWTSPVPLCVCLITCGFAEAGKLLVPMDSGSHMTNOSY
 VEKILRGHEVYVYVPEVSMOLGSLNCTVKTSTYTLDDREFPFAHAKMAY
 RSLFSLTSSNGFNLFFSHCRSLFNDKKEVLEKESFDVFLDPDAGLIVAKY
 FSLPSVFRAGILCHYLEGACCPAPLSVPRILLGSDAMFKERVNNHMLIEHL
 FCGYSKNALEISEILQTPYVDYLSHSISMLRTDVPDYDKRYVNMFTIGIN
 CHGKPLPMEFEATINASGEHGIYVFSLSGSMSEIPEKKAMAIDALGKIPOYVMRY
 TGRPSNLANTTILVKNLPONDILGHPTMRAITHAGSHVETISICNGVPMWPLFG
 DOMNAKMETKAGVTLVNLEMTSEDLNALKAVINDSKYENIMRLSSLHKDRPE
 PLDAVFWVEFVRHKGAPHLRPAHDLTWYQYHSLDVGFLAVLVAVTFAITFKCA
 YGRCGLGKGRVKAKHKSXT"
 109467..>187313
 /gene="UGT1A6"
 109467..109473
 /gene="UGT1A6"
 join(109610..110470,181788..181919,182603..182690,
 182974..183193,187016..>187313)
 /gene="UGT1A6"
 /product="UDP glucuronosyltransferase 1A6"
 join(109610..110470,181788..181919,182603..182690,
 182974..183193,187016..187313)
 /gene="UGT1A6"
 /codon_start=1
 /product="UDP glucuronosyltransferase 1A6"
 /protein_id="AAG30420.1"
 /db_xref="GI:11118745"
 /translation="MACILRSFORISAGVEFLAMGVYDGLVPMDSGSHMTNOSY
 IVEVTSDRGHEVYVYVPEVSMOLGSLNCTVKTSTYTLDDREFPFAHAKMAY
 SFLRPOETETRNMTIVTGLFNCOSLQDRLNFKSKRSDALFDPALPGVLLA
 EYLGIPSVLYRGFPCSLERTSPDVSITPRCTKSDMFTSDVPMWPLVLE
 PYLFYCLFSKYEELASAVLARDVDITLVYKOSWILRLKDFLEFPRVNMWVLE
 INCKRRDLOEFEEAYINASGEHGIYVFSLSGSMSEIPEKKAMAIDALGKIPOYVMRY
 TGRPSNLANTTILVKNLPONDILGHPTMRAITHAGSHVETISICNGVPMWPLFG
 FGDONAKMETKAGVTLVNLEMTSEDLNALKAVINDSKYENIMRLSSLHKDRPE
 PLDAVFWVEFVRHKGAPHLRPAHDLTWYQYHSLDVGFLAVLVAVTFAITFKC
 CAYGRCLGKGRVKAKHKSXT"
 129525..129531
 /gene="UGT1A5"
 129525..>187313
 /gene="UGT1A5"
 join(129600..130466,181788..181919,182603..182690,
 182974..183193,187016..>187313)
 /gene="UGT1A5"
 /product="UDP glucuronosyltransferase 1A5"
 join(129600..130466,181788..181919,182603..182690,
 182974..183193,187016..187313)

Query Match
Best Local Similarity 100.0%; Pred. No. 0.83; Length 198872;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccactggatcacagatct 21
DB 175148 CCACGGGATCAACAGATCT 175128

RESULT 12
LOCUS AY029169 1602 bp mRNA PRI 16-Apr-2001
DEFINITION Macaca mulatta UDP-glucuronosyltransferase UGT1A01 mRNA, complete cds.
ACCESSION AY029169
VERSION AY029169.1 GI:13641264
KEYWORDS
SOURCE
ORGANISM
Thesius monkey.
Macaca mulatta
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecoidea; Macaca.
REFERENCE
AUTHORS Dean, B.J., Zhao, S. and King, C.
TITLE Nucleotide and Deduced Amino Acid Sequence of a Novel UDP-Glucuronosyltransferase 1A Isoform Cloned from Male Rhesus Monkey Liver
JOURNAL unpublished
REFERENCE 2 (bases 1 to 1602)
AUTHORS Dean, B.J., Zhao, S. and King, C.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2001) Drug Metabolism, Merck & Co. Inc., Mail Drop RY80E-109, PO Box 2000, Rahway, NJ 07065, USA

FEATURES
source
1..1602
/organism="Macaca mulatta"
/db_xref="taxon:9544"
/sex="male"
1..1602
/product="UDP-glucuronosyltransferase UGT1A01"
/protein_id="AAK31597.1"
/db_xref="GI:13641265"
/translation="MAVESGGRHPLVGLLCVLPVCHAGKMLIPVDSHMLSM
GTIQLQGRHEIIVLADASYIRREGAFYTLKTPVFPQREDEYSEFVSLGHVFN
DSFLQRIKTKYKIRKDSAMLSGSHLHKELMASLASSFVMDPLPCGPV
AQYSLTPVFFINALPCSLSEBACQCPSPVYPRPISASDHMTFLQKSLTAFS
ONFLCDVYSPTATLSEFLQREYVQNLSSAWLLRSFYDRIPIKPMATIG
GINCLHOSPLSDFEAYINASGHHIVSLGSVWALPEKKAAMIAIALGKIPQV
WRYGTGPSNLANTILVKMLPONDILGHPTAFITAGSHGIEGICNGVPMWP
LFGDQDNKAKRMETKAGMTLVLENTSEDEENLAKAVINDSKYENIMHLSLHKR
PVEPDLAVFVFEVFMHKGAPHLRPAHDLTWQYHSLDIVIGFLAVLVAFIAFK
CCAYGRKCKGKGRVYKAKSKTH"

BASE COUNT 383 a 417 c 390 g 412 t
ORIGIN

Query Match
Best Local Similarity 92.4%; Score 19.4; DB 91; Length 1602;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccactggatcacagatct 21
DB 106 CCACGGGATCAACAGATCT 86

RESULT 13
LOCUS AF104339 1647 bp mRNA PRI 13-DEC-1999
DEFINITION Macaca fascicularis UDP-glucuronosyltransferase UGT1A01 mRNA, complete cds.
ACCESSION AF104339
VERSION AF104339.1 GI:6537143
KEYWORDS
SOURCE
ORGANISM
cra-b-eating macaque.
Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecoidea; Macaca.
REFERENCE
AUTHORS Albert, C., Vallee, M., Beaudry, G., Belanger, A. and Hum, D.W.
TITLE 1 (bases 1 to 1647)
JOURNAL Direct Submission
Submitted (05-NOV-1998) Molecular Endocrinology, Laval University,
2705 Laurier Boulevard, Sainte-Foy, Que G1V4G2, Canada

FEATURES
source
1..1647
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
18..1639
/codon_start=1
/product="UDP-glucuronosyltransferase UGT1A01"
/protein_id="AAFI5349.1"
/db_xref="GI:6537144"
/translation="MAVESGGRHPLVGLLCVLPVCHAGKMLIPVDSHMLSM
GTIQLQGRHEIIVLADASYIRREGAFYTLKTPVFPQREDEYSEFVSLGHVFN
DSFLQRIKTKYKIRKDSAMLSGSHLHKELMASLASSFVMDPLPCGPV
AQYSLTPVFFINALPCSLSEBACQCPSPVYPRPISASDHMTFLQKSLTAFS
ONFLCDVYSPTATLSEFLQREYVQNLSSAWLLRSFYDRIPIKPMATIG
GINCLHOSPLSDFEAYINASGHHIVSLGSVWALPEKKAAMIAIALGKIPQV
WRYGTGPSNLANTILVKMLPONDILGHPTAFITAGSHGIEGICNGVPMWP
LFGDQDNKAKRMETKAGMTLVLENTSEDEENLAKAVINDSKYENIMHLSLHKR
PVEPDLAVFVFEVFMHKGAPHLRPAHDLTWQYHSLDIVIGFLAVLVAFIAFK
CCAYGRKCKGKGRVYKAKSKTH"

BASE COUNT 393 a 431 c 409 g 414 t
ORIGIN

Query Match
Best Local Similarity 92.4%; Score 19.4; DB 88; Length 1647;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccactggatcacagatct 21
DB 143 CCACGGGATCAACAGATCT 123

RESULT 14
LOCUS AC007905 104480 bp DNA HTG 03-OCT-2000
DEFINITION Homo sapiens chromosome 16p24.3 clone RP4-754F23, WORKING DRAFT
SEQUENCE 35 ordered pieces.
ACCESSION AC007905
VERSION AC007905.2 GI:10280722
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 104480)
AUTHORS Krenmadiotis, G., Gardner, A.E., Callen, D.F. and Sutherland, G.R.

TITLE Large Scale Sequencing of the Chromosome 16 region q24.3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 104480)
AUTHORS Kremmidiotis, G., Gardner, A.E., Callen, D.F. and Sutherland, G.R.
TITLE Direct Submission
JOURNAL Submitted (24-JUN-1999) CytoGenetics & Molecular Genetics, Women's
& Children's Hospital, 72 King William Rd, Adelaide, SA 5006,
Australia

On Sep 23, 2000 this sequence version replaced g1:5174819.

Genome Centre : CytoGenetics & Molecular Genetics
Centre code : CMGCH
Website: <http://www.wch.sa.gov.au/labmed/c/genetics/sequencing.html>

* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 1479: contig of 1479 bp in length
1480 1487: gap of unknown length
1488 6699: contig of 5212 bp in length
6700 6707: gap of unknown length
6708 10905: contig of 4198 bp in length
10906 10912: gap of unknown length
10913 11733: contig of 821 bp in length
11734 11740: gap of unknown length
11741 12602: contig of 862 bp in length
12603 12609: gap of unknown length
12610 13856: contig of 1247 bp in length
13857 13863: gap of unknown length
13864 20018: contig of 6155 bp in length
20019 20025: gap of unknown length
20026 21942: contig of 1917 bp in length
21943 21949: gap of unknown length
21950 23920: contig of 1971 bp in length
23921 23927: gap of unknown length
23928 24860: contig of 933 bp in length
24861 24867: gap of unknown length
24868 29279: contig of 4412 bp in length
29280 29286: gap of unknown length
29287 29772: contig of 486 bp in length
29773 29779: gap of unknown length
29780 32411: contig of 2632 bp in length
32412 32418: gap of unknown length
32419 33145: contig of 727 bp in length
33146 33152: gap of unknown length
33153 34862: contig of 1710 bp in length
34863 34869: gap of unknown length
34870 36643: contig of 1774 bp in length
36644 36650: gap of unknown length
36651 37348: contig of 696 bp in length
37349 37355: gap of unknown length
37356 42174: contig of 4819 bp in length
42175 42181: gap of unknown length
42182 45604: contig of 3423 bp in length
45605 45611: gap of unknown length
45612 48485: contig of 2874 bp in length
48486 48492: gap of unknown length
48493 49845: contig of 1353 bp in length
49846 49852: gap of unknown length
49853 51199: contig of 1347 bp in length
51200 51206: gap of unknown length
51207 51513: contig of 307 bp in length
51514 51520: gap of unknown length
51521 52180: contig of 660 bp in length
52181 52187: gap of unknown length
52188 60026: contig of 7839 bp in length
60027 60031: gap of unknown length

60034 64699: contig of 4666 bp in length
64700 64706: gap of unknown length
64707 69136: contig of 4430 bp in length
69137 69143: gap of unknown length
69144 73347: contig of 4204 bp in length
73348 73354: gap of unknown length
73355 82463: contig of 9108 bp in length
82464 82469: gap of unknown length
82470 84166: contig of 1697 bp in length
84167 84173: gap of unknown length
84174 84822: contig of 649 bp in length
84823 84829: gap of unknown length
84830 99511: contig of 14682 bp in length
99512 99518: gap of unknown length
99519 101087: contig of 1569 bp in length
101088 101094: gap of unknown length
101095 103316: contig of 2222 bp in length
103317 103323: gap of unknown length
103324 104480: contig of 1157 bp in length.
Location/Qualifiers
1. 104480
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16q24.3"
/clone="RP4-754F23"

BASE COUNT 23140 a 28756 c 29631 g 22706 t 247 others
ORIGIN

Query Match 84.8%; Score 17.8; DB 60; Length 104480;
Best Local Similarity 90.5%; Pred. No. 52;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cccctggatcacagctatct 21
|||||
DB 20902 CCACCTGGATCAACATATTT 20922

RESULT 15
HSDJ137K2/c
LOCUS
DEFINITION Human DNA sequence from clone RP1-137K2 on chromosome 6q25.1-25.3.
(contains KIAA1235), ESTs, STSS, GSSs and two putative Cpg Islands,
complete sequence.

ACCESSION AL049820
VERSION AL049820.23 GI:8247261
KEYWORDS HTG; B120; Cpg Island; KIAA1235.
SOURCE Human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 187507)

AUTHORS Sycamore, N.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humanyes@sanger.ac.uk

COMMENT On Jun 4, 2000 this sequence version replaced g1:8018160.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:

Em: EMBL, Sw: SWISSPROT, Tr: TREMBL, Wp: WORMPEP, Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/RGP/chr6>

Rp1-137k2 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pletzer de Jong. For further details see <http://Dacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

This sequence is the entire insert of clone Rp1-137k2. The true right end of clone Rp1-80E10 is at 27741 in this sequence.

FEATURES

source

1. 187507
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q25.1-25.3"
/clone="Rp1-137k2"
/clone.lib="RPCI-1"
1. 230
/note="AluSg repeat: matches 2. 237 of consensus"
complement(1808..2197)
/note="match: GSS: Em:AQ106481"
2678..2782
/note="MER53 repeat: matches 85. 189 of consensus"
complement(3848..4325)
/note="match: GSS: Em:B71907"
5108..5417
/note="AluY repeat: matches 1. 297 of consensus"
6157..6278
/note="MIR repeat: matches 15. 153 of consensus"
6589..6661
/note="L2 repeat: matches 2671. 2741 of consensus"
6663..6727
/note="AluY/FLAM repeat: matches 1. 66 of consensus"
6768..7057
/note="L2 repeat: matches 2356. 2661 of consensus"
7142..7944
/note="match: GSS: Em:AQ747336"
7589..7889
/note="AluSx repeat: matches 1. 305 of consensus"
9589..9699
/note="L2 repeat: matches 2579. 2708 of consensus"
9826..10165
/note="L2 repeat: matches 2080. 2417 of consensus"
10580..10752
/note="AluY repeat: matches 132. 302 of consensus"
10795..10905
/note="AluY repeat: matches 48. 158 of consensus"
10975..11081
/note="AluY repeat: matches 14. 125 of consensus"
11763..11796
/note="L17 copies 2 mer tt 82% conserved"
11797..12023
/note="AluSg/X repeat: matches 86. 312 of consensus"
12067..12348
/note="AluSg repeat: matches 1. 280 of consensus"
14350..14877
/note="match: GSS: Em:AQ197120"
14357..14684
/note="match: GSS: Em:AQ312897"
14623..14676
/note="Alu repeat: matches 249. 302 of consensus"
14677..14870
/note="AluY repeat: matches 1. 302 of consensus"
15034..15291
/note="LIMB3 repeat: matches 5921. 6181 of consensus"
16601..16797
/note="MIR repeat: matches 14. 219 of consensus"
17172..17231
/note="30 copies 2 mer tt 81% conserved"

repeat_region 17174..17225
/note="4 copies 13 mer 84% conserved"
repeat_region 17234..17514
/note="AluSx repeat: matches 1. 283 of consensus"
repeat_region 17515..17564
/note="25 copies 2 mer tt 80% conserved"
repeat_region 17517..17556
/note="8 copies 5 mer ttgt 87% conserved"
repeat_region 18805..18879
/note="LIM4C repeat: matches 1685. 1759 of consensus"
repeat_region 19556..19625
/note="MIR repeat: matches 78. 146 of consensus"
repeat_region 19765..20073
/note="AluSg repeat: matches 1. 310 of consensus"
complement(20235..20678)
/note="match: GSS: Em:AQ65222"
22127..22450
/note="L2 repeat: matches 1579. 1915 of consensus"
23241..23349
/note="MER53 repeat: matches 34. 142 of consensus"
23350..23846
/note="MER6 repeat: matches 1. 865 of consensus"
23847..24019
/note="MER33 repeat: matches 141. 324 of consensus"
24447..24872
/note="match: GSS: Em:AQ526363"
24458..24762
/note="match: GSS: Em:AQ665701"
24739..25035
/note="AluY repeat: matches 1. 291 of consensus"
25108..25407
/note="AluSg repeat: matches 2. 302 of consensus"
25938..26203
/note="L2 repeat: matches 2002. 2292 of consensus"
26263..26352
/note="MIR repeat: matches 91. 192 of consensus"
27823..29153
/note="CpG Island"
evidence=not-experimental
27851..28381
/note="match: GSS: Em:AQ281427"
complement(28203..28600)
/note="match: GSS: Em:AQ37194"
28659..29093
/note="match: GSS: Em:AQ37195"
29028..29118
/note="L7 copies 13 mer 70% conserved"
29721..30213
/note="match: GSS: Em:AQ214419"
29780..30137
/note="match: GSS: Em:AQ09774"
29812..30102
/note="match: GSS: Em:AQ096409"
31375..31688
/note="AluSx repeat: matches 1. 308 of consensus"
32258..32813
/note="LIMB1 repeat: matches 5549. 6053 of consensus"
32823..33107
/note="AluSx repeat: matches 1. 297 of consensus"
33581..33709
/note="FLAM repeat: matches 1. 128 of consensus"
complement(33668..34018)
/note="match: STR: Em:G24299"
34949..34972
/note="12 copies 2 mer tt 95% conserved"
36306..36476
/note="L2 repeat: matches 2571. 2750 of consensus"
36527..36791
/note="L2 repeat: matches 2159. 2428 of consensus"
36931..37354
/note="MIR repeat: matches 57. 504 of consensus"
38691..38991
/note="AluSx repeat: matches 1. 296 of consensus"

```

repeat_region 44547..44716
/note="12 repeat: matches 2009..2202 of consensus"
repeat_region 44907..45038
/note="2 copies 66 mer 96% conserved"
repeat_region 45805..46069
/note="AluYb repeat: matches 1..294 of consensus"
repeat_region 46214..46442
/note="AluYb repeat: matches 83..311 of consensus"
repeat_region 46533..46827
/note="6 copies 59 mer 70% conserved"
repeat_region 46550..46747
/note="3 copies 66 mer 74% conserved"
repeat_region 46574..46825
/note="6 copies 42 mer 66% conserved"
repeat_region 46584..46830
/note="19 copies 13 mer 61% conserved"
repeat_region 46714..46771
/note="29 copies 2 mer to 87% conserved"
repeat_region 46835..47132
/note="AluSc repeat: matches 1..308 of consensus"
misc_feature 47007..47346

```

```

Query Match 84.8%; Score 17.8; DB 93; Length 187507;
Best Local Similarity 90.5%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ccactggatcaacagtatct 21
||||| ||| |||||
DB 118886 CCACGCGCATCAGCAGTATCT 118866

```

```

Search completed: July 25, 2001, 05:16:58
Job time: 9221 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 02:43:17 ; Search time 1290.33 Seconds

(without alignments)
251736 Million cell updates/sec

Title: US-09-142-095-1

Perfect score: 21

Sequence: 1 aaagtaactccctgctacactt 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBml:*
1: gb_da1:*
2: gb_da2:*
3: gb_da3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_cm:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pi1:*
13: gb_pi2:*
14: gb_pi3:*
15: gb_pi4:*
16: em_da1:*
17: em_da2:*
18: em_fun:*
19: em_hcgo_hum:*
20: em_hcgo_inv:*
21: em_hcgo_rtd:*
22: em_hcgo_hum1:*
23: em_hcgo_hum2:*
24: em_hcgo_hum3:*
25: em_hcgo_hum4:*
26: em_hcgo_hum5:*
27: em_hcgo_hum6:*
28: em_hcgo_hum7:*
29: em_hcgo_hum8:*
30: em_hcgo_inv1:*
31: em_hcgo_inv2:*
32: em_hcgo_other:*
33: em_hcgo_rtd:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in1:*
42: em_cm:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pi1:*
48: em_pi2:*
49: em_pi3:*
50: em_pi4:*
51: em_pi5:*
52: em_pi6:*
53: gb_da1:*
54: gb_da2:*
55: gb_da3:*
56: gb_in1:*
57: gb_in2:*
58: gb_in3:*
59: gb_cm:*
60: gb_ov:*
61: gb_pat1:*
62: gb_pat2:*
63: gb_ph:*
64: gb_pi1:*
65: gb_pi2:*
66: gb_pi3:*
67: gb_pi4:*
68: gb_pi5:*
69: gb_pi6:*
70: gb_pi7:*
71: gb_pi8:*
72: gb_pi9:*
73: gb_pi10:*
74: gb_pi11:*
75: gb_pi12:*
76: gb_pi13:*
77: gb_pi14:*
78: gb_pi15:*
79: gb_pi16:*
80: gb_pi17:*
81: gb_pi18:*
82: gb_pi19:*
83: gb_pi20:*
84: gb_pi21:*
85: gb_pi22:*
86: gb_pi23:*
87: gb_pi24:*
88: gb_pi25:*
89: gb_pi26:*
90: gb_pi27:*
91: gb_pi28:*
92: gb_pi29:*
93: gb_pi30:*
94: gb_pi31:*
95: gb_pi32:*
96: gb_pi33:*
97: gb_pi34:*
98: em_da3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	9 A65500	A65500 Sequence 1
2	21	100.0	208	89 AF135471	AF135471 Salimiri b
3	21	100.0	234	89 AF135462	AF135462 Pan penis
4	21	100.0	234	89 AF135463	AF135463 Pan trogl
5	21	100.0	531	89 AF352795	AF352795 Homo sapi
6	21	100.0	541	89 AF180372	AF180372 Homo sapi
7	21	100.0	620	9 A65504	A65504 Sequence 5
8	21	100.0	3341	91 D87674	D87674 Homo sapien

9	21	100.0	56770	86	AC006985	AC006985 Homo sapi
10	21	100.0	176619	70	AC026497	AC026497 Homo sapi
11	21	100.0	198872	89	AF297093	AF297093 Homo sapi
12	18.4	87.6	136868	86	AC008278	AC008278 Homo sapi
13	17.8	84.8	7218	91	AP000293	AP000293 Homo sapi
14	17.8	84.8	46831	77	AC087204	AC087204 Homo sapi
15	17.8	84.8	47972	90	AL391843	AL391843 Homo sapi
16	17.8	84.8	57000	88	AC068278	AC068278 Homo sapi
17	17.8	84.8	70068	77	AC090246	AC090246 Homo sapi
18	17.8	84.8	100000	91	AP000043	AP000043 Homo sapi
19	17.8	84.8	100000	91	AP000111	AP000111 Homo sapi
20	17.8	84.8	100000	91	AP000187	AP000187 Homo sapi
21	17.8	84.8	105733	75	AC074136	AC074136 Homo sapi
22	17.8	84.8	125428	88	AC073898	AC073898 Homo sapi
23	17.8	84.8	134493	65	AC018537	AC018537 Homo sapi
24	17.8	84.8	145540	64	AC015962	AC015962 Homo sapi
25	17.8	84.8	145540	64	AC015962	AC015962 Homo sapi
26	17.8	84.8	153234	70	AC026947	AC026947 Homo sapi
27	17.8	84.8	156486	62	AC011657	AC011657 Homo sapi
28	17.8	84.8	156771	82	AL590923	AL590923 Homo sapi
29	17.8	84.8	165987	66	AC021201	AC021201 Homo sapi
30	17.8	84.8	167195	87	AC009044	AC009044 Homo sapi
31	17.8	84.8	170916	76	AC079414	AC079414 Homo sapi
32	17.8	84.8	173985	76	AC084085	AC084085 Homo sapi
33	17.8	84.8	174788	82	AP000756	AP000756 Homo sapi
34	17.8	84.8	175393	73	AC068412	AC068412 Homo sapi
35	17.8	84.8	175587	77	AC087616	AC087616 Homo sapi
36	17.8	84.8	176107	83	AP002393	AP002393 Homo sapi
37	17.8	84.8	176317	62	AC011468	AC011468 Homo sapi
38	17.8	84.8	182815	75	AC011386	AC011386 Homo sapi
39	17.8	84.8	183279	62	AC079014	AC079014 Homo sapi
40	17.8	84.8	191602	73	AC068920	AC068920 Homo sapi
41	17.8	84.8	340000	91	AP001716	AP001716 Homo sapi
42	17.4	82.9	1073	53	CN8063X	Al411475 T7 end of
43	17	81.0	23605	12	AB020747	AB020747 Arabidops
44	17	81.0	105223	12	AC007399	AC007399 Arabidops
45	17	81.0	156970	74	AC073330	AC073330 Homo sapi

ALIGNMENTS

RESULT 1
 A65500 21 bp DNA
 LOCUS A65500
 DEFINITION Sequence 1 from Patent WO9732042.
 ACCESSION A65500
 VERSION A65500.1 GI:4531235
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 21)
 AUTHORS Burchell,B.
 TITLE DRUG TRIAL ASSAY SYSTEM
 JOURNAL Patent: WO 9732042-A 1 04-SEP-1997;
 UNTV DUNDEE (GB)
 COMMENT Other publication AU 2224197 19970916.
 FEATURES
 source location/qualifiers
 1..21
 /db_xref="taxon:32644"

BASE COUNT 5 a 7 c 3 g 6 t
 ORIGIN
 Query Match 100.0%; Score 21; DB 9; Length 21;
 Best Local Similarity 100.0%; Pred. No. 0.96;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aagtgactccctgactct 21
 1 aagtgactccctgactct 21
 Db 1 AAGTGAACCTCCCTGCTACTCT 21

2	208	bp	DNA	PRI	21-NOV-1999	AF135471
LOCUS	AF135471					
DEFINITION	Salmirol boliviensis UDP-glucuronosyltransferase 1A1 (UGT1A1) gene, promoter region and partial cds.					
ACCESSION	AF135471					
VERSION	AF135471.1					
KEYWORDS	Bolivian squirrel monkey.					
SOURCE	Salmirol boliviensis					
ORGANISM	Bolivian squirrel monkey.					
REFERENCE	1 (bases 1 to 208)					
AUTHORS	Hall,D., Ybaveza,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di Rienzo,A.					
TITLE	Variability at the uridine diphosphate glucuronosyltransferase 1A1 promoter in human populations and primates					
JOURNAL	Pharmacogenetics (1999) In press					
REFERENCE	2 (bases 1 to 208)					
AUTHORS	Ybaveza,G., Hall,D. and Di Rienzo,A.					
TITLE	Direct Submission					
JOURNAL	Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924 E. 57th street, Chicago, IL 60637, USA					
FEATURES	location/Qualifiers					
source	1..208					
mrna	/organism="Salmirol boliviensis"					
gene	/db_xref="taxon:27679"					
cds	/gene="UGT1A1"					
	/product="UDP-glucuronosyltransferase 1A1"					
	/db_xref="GI:6456560"					
	/translation="MAAESRGCGQLVGLLCLVGP"					

BASE COUNT 44 a 52 c 58 g 54 t
 ORIGIN
 Query Match 100.0%; Score 21; DB 89; Length 208;
 Best Local Similarity 100.0%; Pred. No. 0.85;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aagtgactccctgactct 21
 3 AAGTGAACCTCCCTGCTACTCT 23
 Db 3 AAGTGAACCTCCCTGCTACTCT 23
 RESULT 3
 AF135462 234 bp DNA PRI 21-NOV-1999
 LOCUS AF135462
 DEFINITION Pan paniscus UDP-glucuronosyltransferase 1A1 (UGT1A1) gene, promoter region and partial cds.
 ACCESSION AF135462
 VERSION AF135462.1 GI:6456541
 KEYWORDS
 SOURCE pygmy chimpanzee.
 ORGANISM Pan paniscus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
 Hall,D., Ybaveza,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di Rienzo,A.
 TITLE Variability at the uridine diphosphate glucuronosyltransferase 1A1 promoter in human populations and primates
 JOURNAL Pharmacogenetics (1999) In press

REFERENCE 2 (bases 1 to 234)
 AUTHORS Ybazeza,G., Hall,D. and Di Rienzo,A.
 TITLE Direct Submission
 JOURNAL Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924 E. 57th Street, Chicago, IL 60637, USA

FEATURES
 source
 1..234
 /organism="Pan paniscus"
 /db_xref="taxon:9597"
 <148..>234
 /gene="UGT1A1"
 /product="UDP-glucuronosyltransferase 1A1"
 <148..>234
 /gene="UGT1A1"
 148..>234
 /gene="UGT1A1"
 /codon_start=1
 /product="UDP-glucuronosyltransferase 1A1"
 /protein_id="AA09173.1"
 /db_xref="GI:6456542"
 /translation="MAVESGGGRPLVGLLCLVAGPVVSHAK"

BASE COUNT 50 a 53 c 69 g 62 t

ORIGIN

Query Match 100.0%; Score 21; DB 89; Length 234;
 Best Local Similarity 100.0%; Pred. No. 0.84; 0; Indels 0; Gaps 0;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 1 aagtgacctcctgctacctt 21
 |||
 3 AAGTGAACCTCCCTGCTACCT 23

RESULT 4
 AF135463 234 bp DNA PRI 21-NOV-1999
 LOCUS AF135463
 DEFINITION Pan troglodytes UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
 promoter region and partial cds.
 ACCESSION AF135463
 VERSION AF135463.1 GI:6456543
 KEYWORDS
 SOURCE
 ORGANISM
 chimpanzee.
 Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
 1 (bases 1 to 234)
 Hall,D., Ybazeza,G., Destro-Bisoli,G., Petzl-Erler,M.L. and Di
 Rienzo,A.
 TITLE Variability at the uridine diphosphate glucuronosyltransferase 1A1
 promoter in human populations and primates
 JOURNAL Pharmacogenetics (1999) In press
 2 (bases 1 to 234)
 Ybazeza,G., Hall,D. and Di Rienzo,A.
 TITLE Direct Submission
 JOURNAL Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
 E. 57th Street, Chicago, IL 60637, USA

FEATURES
 source
 1..234
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 <148..>234
 /gene="UGT1A1"
 /product="UDP-glucuronosyltransferase 1A1"
 <148..>234
 /gene="UGT1A1"
 148..>234
 /gene="UGT1A1"
 /codon_start=1
 /product="UDP-glucuronosyltransferase 1A1"
 /protein_id="AA09174.1"
 /db_xref="GI:6456544"
 /translation="MAVESGGGRPLVGLLCLVAGPVVSHAK"

BASE COUNT 50 a 53 c 69 g 62 t

ORIGIN

Query Match 100.0%; Score 21; DB 89; Length 234;
 Best Local Similarity 100.0%; Pred. No. 0.84; 0; Indels 0; Gaps 0;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 1 aagtgacctcctgctacctt 21
 |||
 3 AAGTGAACCTCCCTGCTACCT 23

RESULT 5
 AF352795 531 bp DNA PRI 10-APR-2001
 LOCUS AF352795
 DEFINITION Homo sapiens bilirubin UDP-glucuronosyltransferase 1-1 (UGT1A1) gene,
 UGT1A1*1 allele, partial cds.
 ACCESSION AF352795
 VERSION AF352795.1 GI:13569708
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 531)
 McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
 7(TA) repeat polymorphism of the TATA box of human bilirubin
 UDP-glucuronosyltransferase 1-1(UGT1A1*) gene in a patient with
 sickle cell anemia + high bilirubinemia
 Unpublished
 2 (bases 1 to 531)
 McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
 TITLE Direct Submission
 JOURNAL Submitted (23-FEB-2001) Medicine/Hemoglobin DNA Laboratory, Medical
 College of Georgia, 15th St., NC-1000, Augusta, GA 30912, USA

FEATURES
 source
 1..531
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2q37"
 /cell_type="white blood cell"
 /tissue_type="blood"
 <1..>531
 /gene="UGT1A1"
 /note="UGT1"
 /allele="UGT1A1*1"
 <1..313
 /gene="UGT1A1"
 279..295
 /gene="UGT1A1"
 279..292
 /note="repeat polymorphism compared to UGT1A1 sequence
 presented in Genbank Accession Number AF180372; contains 7
 ta repeats frequently detected in patients with high
 bilirubinemia"
 /rpt_type="tandem
 /rpt_unit="ta
 314..>531
 /gene="UGT1A1"
 /product="bilirubin UDP-glucuronosyltransferase 1-1"
 334..>531
 /gene="UGT1A1"
 /codon_start=1
 /product="bilirubin UDP-glucuronosyltransferase 1-1"
 /protein_id="AAK1204.1"
 /db_xref="GI:13569709"
 /translation="MAVESGGGRPLVGLLCLVAGPVVSHAKTLLIPDSSHLNKL
 GATQDQGRSHIVLAPDAST"

BASE COUNT 120 a 121 c 137 g 153 t

ORIGIN

Query Match 100.0%; Score 21; DB 89; Length 541;
 Best Local Similarity 100.0%; Pred. No. 0.81;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aagtgactccctgctacctt 21
 |||
 DB 185 AAGTGAACCTCCTGCTACCTT 205

RESULT 6

LOCUS AF180372 541 bp DNA PRI 05-OCT-1999
 DEFINITION Homo sapiens bilirubin UDP-glucuronosyltransferase 1-1 (UGT1) gene,
 UGT1*1 allele, partial cds.
 ACCESSION AF180372
 VERSION AF180372.1 GI:6010649
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 541)
 Kular, F., Stromek, E., Leitner, C., Nechtman, J. and Kular, A.
 TITLE Detection of the TATA box polymorphism of the human bilirubin
 UDP-glucuronosyltransferase 1-1 gene (UGT1*1) in a patient with
 sickle cell anemia

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 541)
 Kular, F., Stromek, E., Leitner, C., Nechtman, J. and Kular, A.
 AUTHORS Direct Submission
 TITLE Submitted (24-AUG-1999) Medicine, Hematology/Oncology-Sickle Cell
 JOURNAL Center, Medical College of Georgia, 15th Street, Ac-1000, Augusta,
 GA 30912, USA

FEATURES
 source Location/Qualifiers
 1..541
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2q37"
 /cell_type="white blood cells"
 /tissue_type="whole blood"
 1..>541
 /gene="UGT1"
 /note="UGT1, UGT1A1"
 /allele="UGT1*1"
 1..313
 /gene="UGT1"
 226
 /gene="UGT1"
 /replaces="c"
 279..288
 /note="polymorphic region"
 /rpl_type=" tandem
 279..291
 /rpl_type=" tandem
 TATA_signal
 /note="Ritter, J.K., et al., 1992, J. Biol. Chem.,
 267:3257-3261"
 314..>541
 /gene="UGT1"
 /product="bilirubin UDP-glucuronosyltransferase 1-1"
 330..>541
 /gene="UGT1"
 /note="UGT1"
 /note="UDP glucosyltransferase 1"
 /codon_start=1
 /product="bilirubin UDP-glucuronosyltransferase 1-1"
 /protein_id="AF01205.1"
 /db_xref="GI:6010650"
 /translation="MAVESQGRPRVYGLLCVIGPVVSHACKILLIFVDSHSLMT
 GAIIDQQRGHEIVYLAIPASLIRDG"

BASE COUNT 123 a 124 c 141 g 153 t
 ORIGIN

Query Match 100.0%; Score 21; DB 89; Length 541;
 Best Local Similarity 100.0%; Pred. No. 0.81;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aagtgactccctgctacctt 21
 |||
 DB 185 AAGTGAACCTCCTGCTACCTT 205

RESULT 7

LOCUS A65504 620 bp DNA PAT 29-MAR-1999
 DEFINITION Sequence 5 from Patent WO9732042.
 ACCESSION A65504
 VERSION A65504.1 GI:4531239
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 unclassified.

REFERENCE 1 (bases 1 to 620)
 Burchell, B.
 AUTHORS DRUG TRIAL ASSAY SYSTEM
 TITLE Patent: WO 9732042-A 5 04-SEP-1997;
 JOURNAL UNIV DUNDEE (GB)
 COMMENT Other publication AU 2224197 19970916.
 FEATURES
 source Location/Qualifiers
 1..620
 /organism="unidentified"
 /db_xref="taxon:32644"

BASE COUNT 157 a 127 c 151 g 185 t
 ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 620;
 Best Local Similarity 100.0%; Pred. No. 0.8;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aagtgactccctgctacctt 21
 |||
 DB 465 AAGTGAACCTCCTGCTACCTT 485

RESULT 8

LOCUS D87674 3341 bp DNA PRI 14-APR-2000
 DEFINITION Homo sapiens gene for bilirubin UDP-glucuronosyltransferase 1,
 promoter region and partial cds.
 ACCESSION D87674
 VERSION D87674.1 GI:3059176
 KEYWORDS bilirubin UDP-glucuronosyltransferase 1.
 SOURCE Homo sapiens DNA.
 ORGANISM Homo sapiens

REFERENCE 1 (sites)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE Analysis of the promoter of human bilirubin
 UDP-glucuronosyltransferase gene (UGT1*1) in relevance to Gilbert's
 syndrome

HEPATOL. Res. 9, 152-163 (1997)
 2 (bases 1 to 3341)
 /note="UGT1"
 /note="UDP glucosyltransferase 1"
 /codon_start=1
 /product="bilirubin UDP-glucuronosyltransferase 1-1"
 /protein_id="AF01205.1"
 /db_xref="GI:6010650"
 /translation="MAVESQGRPRVYGLLCVIGPVVSHACKILLIFVDSHSLMT
 GAIIDQQRGHEIVYLAIPASLIRDG"

COMMENT Sequence updated (08-Jan-1997) by: Hisao Ueyama.
 FEATURES
 source Location/Qualifiers
 1..3341


```

repeat_region      /rpt_family="MERL_type"
                    4683..4806
                    /rpt_family="MIR"
repeat_region      /rpt_family="MIR"
                    5131..5389
                    /rpt_family="MALR"
repeat_region      8738..8871
                    /rpt_family="MIR"
repeat_region      9092..9371
                    /rpt_family="Achoho"
repeat_region      9967..10582
                    /rpt_family="L1"
repeat_region      10576..11200
                    /rpt_family="L1"
repeat_region      11196..11469
                    /rpt_family="L1"
repeat_region      11540..11654
                    /rpt_family="L1"
repeat_region      12108..12523
                    /rpt_family="MALR"
repeat_region      13113..13149
                    /rpt_family="MERL_type"
repeat_region      13150..13171
                    /rpt_family="CAG)n"
repeat_region      13172..13286
                    /rpt_family="MERL_type"
repeat_region      13338..13497
                    /rpt_family="MERL_type"
repeat_region      14727..14848
                    /rpt_family="MIR"
repeat_region      15890..16244
                    /rpt_family="MALR"
repeat_region      16380..16723
                    /rpt_family="L1"
repeat_region      17483..17560
                    /rpt_family="L2"
repeat_region      17655..17747
                    /rpt_family="L2"
repeat_region      18220..18337
                    /rpt_family="MERL_type"
repeat_region      18770..18816
                    /rpt_family="CA)n"
repeat_region      19103..19121
                    /rpt_family="A-rich"
repeat_region      19122..19636
                    /rpt_family="L1"
repeat_region      19637..19662
                    /rpt_family="A-rich"
repeat_region      19825..20014
                    /rpt_family="L1"
repeat_region      20034..20098
                    /rpt_family="MIR"
repeat_region      20254..21067
                    /rpt_family="L1"
repeat_region      21411..21182
                    /rpt_family="A-rich"
repeat_region      24649..25018
                    /rpt_family="Retroviral"
repeat_region      25258..25447
                    /rpt_family="MIR"
repeat_region      25777..25824
                    /rpt_family="CA)n"
repeat_region      28020..28178
                    /rpt_family="MIR"
repeat_region      28178..28222
                    /rpt_family="L2"
repeat_region      28503..28558
                    /rpt_family="L1"
repeat_region      28358..28741
                    /rpt_family="L1"
repeat_region      28970..29064
                    /rpt_family="L1"
repeat_region      29276..29434
                    /rpt_family="MERL_type"

```

```

repeat_region      29554..29712
                    /rpt_family="MERL_type"
repeat_region      30867..31015
                    /rpt_family="MERL_type"
repeat_region      31268..31682
                    /rpt_family="MALR"
repeat_region      32195..32390
                    /rpt_family="MIR"
repeat_region      33607..33760
                    /rpt_family="L2"
repeat_region      33835..34127
                    /rpt_family="Alu"
repeat_region      34167..34283
                    /rpt_family="GA-rich"
repeat_region      34806..34976
                    /rpt_family="L2"
repeat_region      35247..35719
                    /rpt_family="L2"
repeat_region      35720..35762
                    /rpt_family="GA-rich"
repeat_region      35763..35909

```

Query Match 100.0%; Score 21; DB 86; Length 68770;
 Best Local Similarity 100.0%; Pred. No. 0.62;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aagtgaaatccctgcctactt 21
 Db 66302 AAGTGAATCCCTGCCTACTT 66282

RESULT 10
 AC026497/C
 LOCUS 176619 bp DNA HTG 22-MAR-2000
 DEFINITION Homo sapiens chromosome 11 clone RP11-689A10 map 11, *** SEQUENCING
 IN PROGRESS ***; 43 unordered pieces.
 AC026497
 VERSION AC026497.1 GI:7283186
 KEYWORDS HTG; HTGS; PHASEL.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 176619)
 Birten, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhalter, B., Brown, A., Burdett, G.,
 Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Gargan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand, P., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Melchior, J., Meneses, L., Mihova, T., Miranda, C., Mogens, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pleasant, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange, Rhomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testa, S., Theodore, J., Tirrell, A., Travers, M., Triggilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission
 JOURNAL Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT All repeats were identified using RepeatMasker:


```
misc_feature      30235..32646
                  /note="assembly_fragment"
misc_feature      32747..35096
                  /note="assembly_fragment"
                  clone_end:r7
                  vector_side:right"
misc_feature      35197..37696
                  /note="assembly_fragment"
misc_feature      37797..40539
                  /note="assembly_fragment"
misc_feature      40640..44518
                  /note="assembly_fragment"
misc_feature      45619..48728
                  /note="assembly_fragment"

Query Match      100.0% Score 21; DB 70; Length 176619;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 aagtaactcctgctactctt 21
          |||
Db      89548 AAGTGAAGCTCCTGCTACTCTT 89528

RESULT 11
LOCUS      AF297093      198872 bp      DNA      PRI      07-NOV-2000
DEFINITION Homo sapiens UGT1 gene locus, complete sequence.
ACCESSION   AF297093
VERSION     AF297093.1 GI:11118740
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 198872)
            Owens,I.S., Gong,Q., Cho,J.W., Huang,T., Potter,C., Gholami,N.,
            Basu,N.K., Kubota,S., Carvalho,S. and Pennington,M.W.
            Thirteen UDP glucuronosyltransferase genes encoded at the human
            UGT1 locus
            Unpublished
            2 (bases 1 to 198872)
            Owens,I.S., Gong,Q., Cho,J.W., Huang,T., Potter,C., Gholami,N.,
            Basu,N.K., Kubota,S., Carvalho,S. and Pennington,M.W.
            Direct Submision
            Submitted (10-AUG-2000) Heritable Disorders Branch, NIH-NICHD, 9000
            Rockville Pike, Bethesda, MD 20892, USA
FEATURES
            source
            1..198872
            Location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1989..2841
            /gene="UGT1A12p"
            /note="UDP glucuronosyltransferase 1A12"
            /pseudo
            1989..2841
            /gene="UGT1A12p"
            /note="UDP glucuronosyltransferase 1A12"
            /pseudo
            19970..20956
            /gene="UGT1A11p"
            /note="UDP glucuronosyltransferase 1A11"
            /pseudo
            19970..19978
            /gene="UGT1A11p"
            /note="UDP glucuronosyltransferase 1A11"
            /pseudo
            20105..20956
            /gene="UGT1A11p"
            /note="UDP glucuronosyltransferase 1A11"
            /pseudo
            /codon_start=1
            /pseudo
            34133..34149
            /gene="UGT1A8"
            /note="UGT1A8"
            34133..>187313
            /gene="UGT1A8"
            /note="UGT1A8"
            join(<34264..35118,181788..181919,182603..182690,

182974..183193,187016..>187313)
/gene="UGT1A8"
/product="UDP glucuronosyltransferase 1A8"
join(<34264..35118,181788..181919,182603..182690,
182974..183193,187016..187313)
/gene="UGT1A8"
/codon_start=1
/product="UDP glucuronosyltransferase 1A8"
/protein_id="AAG30416.1"
/db_xref="GI:11118741"
/translation="MARGMTSPPLPCVILLTCGFAEAGKLLVPMDSHMTMOSV
VEKILLGHEVYVYVMPYNSKOLGSLNCTVKTSTSTYEDIDREPDADPAQAKAOV
RSLPSLFLSSNGFENLFEPSRCLFNDKRLVLEESFDVAVLPDPDCCGLYAKY
FSLPSVYFANGIACHYLEGACAPPLSYVRIILLESFDMATFERYRNHIMLEBHL
FCOYFSKNALEIASLEILQPTVAYDLVSHRSIMLEPTDVLDPYRPMVPMITIGIN
CHQGRPLDMEFEAVINASGEHIVFSLGSAVSEIPEKAKMAIDALGKIPQVLMRY
TGTRPSNLANTITLVKMLPONDLGHPTRAFTFHAGSGVYSEICNGVPMWAPLFG
DOMDNARMETKAGATYTLNLEMTSEDEENAKAVINDKSYSEKNTMRLSLHDPREY
PLDLAVFVFEVVRHKGAPHLRPAHDLTWQYHSLDVGFLAVLVAVATFEKCA
YGRKCLGKGRVKAHKSRTH"
52972..52988
/gene="UGT1A10"
52972..>187313
/gene="UGT1A10"
join(<53103..53957,181788..181919,182603..182690,
182974..183193,187016..>187313)
/gene="UGT1A10"
/product="UDP glucuronosyltransferase 1A10"
join(<53103..53957,181788..181919,182603..182690,
182974..183193,187016..187313)
/gene="UGT1A10"
/codon_start=1
/product="UDP glucuronosyltransferase 1A10"
/protein_id="AAG30417.1"
/db_xref="GI:11118742"
/translation="MARAGMTSPVPLCVILLTCGFAEAGKLLVPMDSHMTMOSV
VEKILLGHEVYVYVMPYNSKOLGSLNCTVKTSTSTYEDIDREPDADPAQAKAOV
OISFSLSSSSGGLDFTFSRCSLFNDKRLVLEESFDVAVLPDPDCCGLYAKY
FSLPSVYFANGIACHYLEGACAPPLSYVRIILLESFDMATFERYRNHIMLEBHL
FCOYFSKNALEIASLEILQPTVAYDLVSHRSIMLEPTDVLDPYRPMVPMITIGIN
CHQGRPLDMEFEAVINASGEHIVFSLGSAVSEIPEKAKMAIDALGKIPQVLMRY
TGTRPSNLANTITLVKMLPONDLGHPTRAFTFHAGSGVYSEICNGVPMWAPLFG
DOMDNARMETKAGATYTLNLEMTSEDEENAKAVINDKSYSEKNTMRLSLHDPREY
PLDLAVFVFEVVRHKGAPHLRPAHDLTWQYHSLDVGFLAVLVAVATFEKCA
YGRKCLGKGRVKAHKSRTH"
64509..65608
/gene="UGT1A13p"
/note="UDP glucuronosyltransferase 1A13"
/pseudo
join(<64509..65011,65353..65608)
/gene="UGT1A13p"
/codon_start=1
/pseudo
88416..>187313
/gene="UGT1A9"
88416..88427
/gene="UGT1A9"
join(<88543..89397,181788..181919,182603..182690,
182974..183193,187016..>187313)
/gene="UGT1A9"
/product="UDP glucuronosyltransferase 1A9"
join(<88543..89397,181788..181919,182603..182690,
182974..183193,187016..187313)
/gene="UGT1A9"
/codon_start=1
/product="UDP glucuronosyltransferase 1A9"
/protein_id="AAG30418.1"
/db_xref="GI:11118743"
/translation="MACTGWTSPPLPCVILLTCGFAEAGKLLVPMDSHMTMOSV
VEKILLGHEVYVYVMPYNSKOLGSLNCTVKTSTSTYEDIDREPDADPAQAKAOV
RSLPSLFLSSNGFENLFEPSRCLFNDKRLVLEESFDVAVLPDPDCCGLYAKY
FSLPSVYFANGIACHYLEGACAPPLSYVRIILLESFDMATFERYRNHIMLEBHL
FCOYFSKNALEIASLEILQPTVAYDLVSHRSIMLEPTDVLDPYRPMVPMITIGIN
CHQGRPLDMEFEAVINASGEHIVFSLGSAVSEIPEKAKMAIDALGKIPQVLMRY
TGTRPSNLANTITLVKMLPONDLGHPTRAFTFHAGSGVYSEICNGVPMWAPLFG
DOMDNARMETKAGATYTLNLEMTSEDEENAKAVINDKSYSEKNTMRLSLHDPREY
PLDLAVFVFEVVRHKGAPHLRPAHDLTWQYHSLDVGFLAVLVAVATFEKCA
YGRKCLGKGRVKAHKSRTH"
LCHRFKNALEIASLEILQPTVETDLYSHTSIMLRTDVLDPYRPMVPMITIGIN
```

```

CHGKRLPMEFEAYINASGEHIVTSLSMSSELPKKAALADALGKIPOTVLMRY
TGRPSNLANTLIVKMLPONDLLGHMPRAFTHAGSHGYTESICGVPMMPLFG
DQDNNAKRETKGAGVTLNVLNEMSEDLNAKAVINDSKYKEMIMLSLKHDRVE
PLDAVFEVPMRKGAFLRPAADLTMVYHSLDVLIGLLAVLVAFTFPCCA
YGRKCLGKKGRVKKAKHAKSKTH"
98424..98438
/gene="UGT1A7"
98424..>187313
/join(<98452..99406,181788..181919,182603..182690,
182974..183193,187016..>187313)
/gene="UGT1A7"
/product="UDP glucuronosyltransferase 1A7"
/join(98452..99406,181788..181919,182603..182690,
182974..183193,187016..187313)
/gene="UGT1A7"
/codon_start=1
/product="UDP glucuronosyltransferase 1A7"
/protein_id="AAG30419.1"
/db_xref="GI:1118744"
/translation="MARAGMTGLIPYVCLLTTCGFARAGKLVVPMDSHMTQSV
VEKLLRGEHVYVMEVSMQGLNCTVKTYSYSLDEQDEEYFADARMTAPL
RSPASLTSSNGIFDLFFSNCRSTFKDKLYEYLKESCPAVFLDPACGLVAKY
FSLPSTVPRKGRFCHYLEEGNCRAPLAVSPRLIGPSDAPREKRYRNHLEHIL
FCYFFKNVLEIASLIDTPTVATDLSTSTWLLRTPTVLEPKPMNMLPTGGIN
CHGKRPVMEFEAYINASGEHIVTSLSMSSELPKKAALADALGKIPOTVLMRY
TGRPSNLANTLIVKMLPONDLLGHMPRAFTHAGSHGYTESICGVPMMPLFG
DQDNNAKRETKGAGVTLNVLNEMSEDLNAKAVINDSKYKEMIMLSLKHDRVE
PLDAVFEVPMRKGAFLRPAADLTMVYHSLDVLIGLLAVLVAFTFPCCA
YGRKCLGKKGRVKKAKHAKSKTH"
109467..>187313
/gene="UGT1A6"
109467..109473
/join(<109610..110470,181788..181919,182603..182690,
182974..183193,187016..>187313)
/gene="UGT1A6"
/product="UDP glucuronosyltransferase 1A6"
/join(109610..110470,181788..181919,182603..182690,
182974..183193,187016..187313)
/gene="UGT1A6"
/codon_start=1
/product="UDP glucuronosyltransferase 1A6"
/protein_id="AAG30420.1"
/db_xref="GI:1118745"
/translation="MACLRSFORISAGVFALMGVNGDKLLVFPQDSHMLSKMD
IVYLSDRGHEIVVPEVNLLEKSYTRKIVYVPODELNKRYOSGNHFAER
SFLAPOTERYNNMVLIGLYFINGOSLDRDNLNFEKESFDALFTDPLPGVTLA
EYLGLSVYLRGPRCSLEHFSRSDPVSYPDRTYKFSHMFESORVAFNPLVLE
PIYFPLRSKIELASANTLRDNDITLLOKYSVLLAYDVLVPRVPMNVPVIGG
INCKRRKDLDSQEFAYINASGEHIVTSLSMSSELPKKAALADALGKIPOTVLM
RYTGRPSNLANTLIVKMLPONDLLGHMPRAFTHAGSHGYTESICGVPMMPL
FGDQDNNAKRETKGAGVTLNVLNEMSEDLNAKAVINDSKYKEMIMLSLKHDR
VEPLDAVFEVPMRKGAFLRPAADLTMVYHSLDVLIGLLAVLVAFTFPCCA
YGRKCLGKKGRVKKAKHAKSKTH"
129525..>187313
/gene="UGT1A5"
129525..129531
/join(<129600..130466,181788..181919,182603..182690,
182974..183193,187016..>187313)
/gene="UGT1A5"
/product="UDP glucuronosyltransferase 1A5"
/join(129600..130466,181788..181919,182603..182690,
182974..183193,187016..187313)
/gene="UGT1A5"
/codon_start=1
/product="UDP glucuronosyltransferase 1A5"
/protein_id="AAG30421.1"
/db_xref="GI:1118746"
/translation="MANGIAPLPQALTGILLLVQPMASGKVLVVPDGSNMLSM
REALDLRAGHGVVTLNVLNEMSEDLNAKAVINDSKYKEMIMLSLKHDRVE
PLDAVFEVPMRKGAFLRPAADLTMVYHSLDVLIGLLAVLVAFTFPCCA
YGRKCLGKKGRVKKAKHAKSKTH"

```

```

Query Match      100.0%; Score 21; DB 89; Length 198872;
Best Local Similarity 100.0%; Pred. No. 0.59
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 aagtgaaactccctgactt 21
|||||
Db 174896 AAGTGAACTCCCTGACTCTT 174916

```

RESULT 12

AC008278/c

AC008278 136868 bp DNA PRI 07-OCT-2000

DEFINITION Homo sapiens BAC clone RP11-422A6 from 2, complete sequence.

AC008278

AC008278.3 GI:7408138

HUG

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

2 (bases 1 to 136868)

Du, F., Stoneking, T. and Doeber, A.

The sequence of Homo sapiens BAC clone RP11-422A6

unpublished

3 (bases 1 to 136868)

Waterson, R.H.

Direct Submission

Submitted (31-JUL-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

4 (bases 1 to 136868)

Waterson, R.H.

Direct Submission

Submitted (07-APR-2000) Genome Sequencing Center, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Apr 4, 2000 this sequence version replaced gi.5870295.

Genome Center

Center: Washington University, Genome Sequencing Center

Center code: WUSC

Web site: http://genome.wustl.edu/gsc

Contact: saplens@wustl.edu

Summary Statistics

Center project name: R_NH0422A06

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Frengen, E.,
Tateno, M., Catanesse, J. J., and de Jong, P. J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Flieter de Jong
(<http://bacpac.med.buffalo.edu>)
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-32P22, 200 bp overlap; the
clone sequenced to the right is RP11-123E16. Actual start of this
clone is at base position 149633 of RP11-32P22; actual end is at
base position 136868 of RP11-422A6.

FEATURES

source

```

1. 136868
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="2"
   /map="2"
   /clone="RP11-422A6"
   /clone_1lb="RPCI-11"
   1. 70
   /rpt_family="L2"
   533. 1406
   /rpt_family="Retroviral"
   /rpt_family="Retroviral"
   1407. 1441
   /rpt_family="(TAA)n"
   1442. 1453
   /rpt_family="Retroviral"
   2009. 2244
   /rpt_family="L2"
   2245. 2618
   /rpt_family="MALR"
   2619. 2749
   /rpt_family="L2"
   3143. 3211
   /rpt_family="GA-rich"
   3254. 3515
   /rpt_family="L1"
   3516. 3808
   /rpt_family="Alu"
   3824. 4022
   /rpt_family="L1"
   4018. 4666
   /rpt_family="L1"
   5232. 5530
   /rpt_family="Alu"
   5444. 5782
   /note="match to EST AA610222 (NID:g2458650) np90c01.s1"
   7003. 7144
   /rpt_family="L2"
   7145. 7433
   /rpt_family="Alu"
   7434. 7478
   repeat_region
   /rpt_family="L2"

```

```

repeat_region 7840. 8038
   /rpt_family="MIR"
repeat_region 8232. 8512
   /rpt_family="Alu"
repeat_region 8513. 8556
   /rpt_family="(TAAA)n"
repeat_region 8656. 8669
   /rpt_family="Alu"
repeat_region 9127. 9203
   /rpt_family="MIR"
repeat_region 9361. 9669
   /rpt_family="Alu"
repeat_region 9812. 9922
   /rpt_family="L2"
repeat_region 10144. 10164
   /rpt_family="AT-rich"
repeat_region 10591. 10670
   /rpt_family="MER1-type"
repeat_region 10842. 10870
   /rpt_family="AT-rich"
repeat_region 11108. 11335
   /rpt_family="MER2-type"
repeat_region 11579. 11856
   /rpt_family="Alu"
repeat_region 12329. 12363
   /rpt_family="AT-rich"
repeat_region 12469. 12692
   /rpt_family="CT-rich"
repeat_region 12848. 12923
   /note="match to EST AA247962 (NID:g1880769)"
repeat_region 13890. 14062
   /rpt_family="Alu"
repeat_region 14856. 14900
   /rpt_family="(CAT)n"
repeat_region 15368. 15604
   /note="match to EST AA247962 (NID:g1880769)"
repeat_region 16857. 16916
   /rpt_family="MER1-type"
repeat_region 16916. 17114
   /rpt_family="MER1-type"
repeat_region 17199. 17253
   /rpt_family="(TA)n"
repeat_region 17365. 17419
   /note="match to EST AA247962 (NID:g1880769)"
repeat_region 17575. 17872
   /rpt_family="Alu"
repeat_region 18341. 18450
   /rpt_family="L2"
repeat_region 18950. 19018
   /rpt_family="L2"
repeat_region 19030. 19063
   /rpt_family="(TAAA)n"
repeat_region 19260. 19380
   /rpt_family="L2"
repeat_region 20041. 20093
   /note="match to EST AA247962 (NID:g1880769)"
repeat_region 20254. 20359
   /rpt_family="MIR"
repeat_region 20461. 20637
   /rpt_family="L2"
repeat_region 20665. 21076
   /rpt_family="MER4-group"
repeat_region 21076. 21188
   /rpt_family="MER4-group"
repeat_region 21288. 21313
   /rpt_family="(TTCA)n"
repeat_region 21490. 22051
   /rpt_family="L2"
repeat_region 22592. 22833
   /rpt_family="Alu"
repeat_region 22955. 23059
   /note="match to EST A1150474 (NID:g3678943) qf41d07.x1"
repeat_region 23322. 23450

```


RESULT	14
AC087204	
LOCUS	AC087204 46831 bp DNA
DEFINITION	Homo sapiens chromosome 8 clone RP11-520F7 map 8, LOW-PASS SEQUENCING.
ACCESSION	AC087204
VERSION	AC087204.1 GI:11693386
KEYWORDS	HTG; HTGS_PHASE0.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

* NOTE: This record contains 61 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 676: contig of 676 bp in length
*
* 677 776: gap of 100 bp
*
* 777 1441: contig of 665 bp in length
*
* 1442 1541: gap of 100 bp
*
* 1542 2219: contig of 678 bp in length
*
* 2220 2319: gap of 100 bp
*
* 2320 2980: contig of 661 bp in length
*
* 2981 3080: gap of 100 bp
*
* 3081 3757: contig of 677 bp in length
*
* 3758 3857: gap of 100 bp
*
* 3858 4517: contig of 660 bp in length
*
* 4518 4617: gap of 100 bp
*
* 4618 5331: contig of 714 bp in length
*
* 5332 5431: gap of 100 bp
*
* 5432 6074: contig of 643 bp in length
*
* 6075 6174: gap of 100 bp
*
* 6175 6852: contig of 678 bp in length
*
* 6853 6952: gap of 100 bp
*
* 6953 7610: contig of 658 bp in length
*
* 7611 7710: gap of 100 bp

```

```

* 7711 8355: contig of 645 bp in length
* 8356 8455: gap of 100 bp
* 8456 9129: contig of 674 bp in length
* 9130 9223: gap of 100 bp
* 9223 9891: contig of 662 bp in length
* 9892 9991: gap of 100 bp
* 9992 10665: contig of 674 bp in length
* 10666 10765: gap of 100 bp
* 10766 11423: contig of 658 bp in length
* 11424 11523: gap of 100 bp
* 11524 12180: contig of 657 bp in length
* 12181 12280: gap of 100 bp
* 12281 12945: contig of 665 bp in length
* 12946 13045: gap of 100 bp
* 13046 13692: contig of 647 bp in length
* 13693 13792: gap of 100 bp
* 13793 14442: contig of 650 bp in length
* 14443 14542: gap of 100 bp
* 14543 15203: contig of 661 bp in length
* 15204 15303: gap of 100 bp
* 15304 15968: contig of 665 bp in length
* 15969 16068: gap of 100 bp
* 16069 16747: contig of 679 bp in length
* 16748 16847: gap of 100 bp
* 16848 17509: contig of 662 bp in length
* 17510 17609: gap of 100 bp
* 17610 18256: contig of 647 bp in length
* 18257 18356: gap of 100 bp
* 18357 19030: contig of 674 bp in length
* 19031 19130: gap of 100 bp
* 19131 19779: contig of 649 bp in length
* 19780 19879: gap of 100 bp
* 19880 20538: contig of 659 bp in length
* 20539 20638: gap of 100 bp
* 20639 21294: contig of 656 bp in length
* 21295 21394: gap of 100 bp
* 21395 22069: contig of 675 bp in length
* 22070 22169: gap of 100 bp
* 22170 22845: contig of 676 bp in length
* 22846 22945: gap of 100 bp
* 22946 23627: contig of 682 bp in length
* 23628 23727: gap of 100 bp
* 23728 24395: contig of 668 bp in length
* 24396 24495: gap of 100 bp
* 24496 25150: contig of 655 bp in length
* 25151 25250: gap of 100 bp
* 25251 25913: contig of 663 bp in length
* 25914 26013: gap of 100 bp
* 26014 26661: contig of 648 bp in length
* 26662 26761: gap of 100 bp
* 26762 27441: contig of 680 bp in length
* 27442 27541: gap of 100 bp
* 27542 28201: contig of 660 bp in length
* 28202 28301: gap of 100 bp
* 28302 28991: contig of 680 bp in length
* 28992 29091: gap of 100 bp
* 29092 29766: contig of 675 bp in length
* 29767 29866: gap of 100 bp
* 29867 30521: contig of 655 bp in length
* 30522 30621: gap of 100 bp
* 30622 31300: contig of 679 bp in length
* 31301 31400: gap of 100 bp
* 31401 32074: contig of 674 bp in length
* 32075 32174: gap of 100 bp
* 32175 32854: contig of 680 bp in length
* 32855 32954: gap of 100 bp
* 32955 33638: contig of 685 bp in length
* 33640 33739: gap of 100 bp
* 33740 34430: contig of 691 bp in length
* 34431 34530: gap of 100 bp
* 34531 35185: contig of 655 bp in length
* 35186 35285: gap of 100 bp
* 35286 35958: contig of 673 bp in length

```

FEATURES

source

```

* 35959 36058: gap of 100 bp
* 36059 36729: contig of 671 bp in length
* 36730 36829: gap of 100 bp
* 36830 37509: contig of 680 bp in length
* 37510 37609: gap of 100 bp
* 37610 38263: contig of 654 bp in length
* 38264 38363: gap of 100 bp
* 38364 39039: contig of 676 bp in length
* 39040 39139: gap of 100 bp
* 39140 39807: contig of 668 bp in length
* 39808 39907: gap of 100 bp
* 39908 40578: contig of 671 bp in length
* 40579 40678: gap of 100 bp
* 40679 41357: contig of 679 bp in length
* 41358 41457: gap of 100 bp
* 41458 42167: contig of 710 bp in length
* 42168 42267: gap of 100 bp
* 42268 42963: contig of 696 bp in length
* 42964 43063: gap of 100 bp
* 43064 43713: contig of 650 bp in length
* 43714 43813: gap of 100 bp
* 43814 44501: contig of 688 bp in length
* 44502 44601: gap of 100 bp
* 44602 45262: contig of 661 bp in length
* 45263 45362: gap of 100 bp
* 45363 46040: contig of 678 bp in length
* 46041 46140: gap of 100 bp
* 46141 46831: contig of 691 bp in length.

```

```

Location/Qualifiers
1. 46831
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-520F7"
/clone_lib="RPCT-11 Human Male BAC"

```

```

BASE COUNT      10780 a      8764 c      9356 g      11744 t      6187 others
ORIGIN

```

Query Match

```

Best Local Similarity 84.8%; Score 17.8; DB 77; Length 46831;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 aagtaactccctcactcctt 21
|||||
Db 7326 AGCTGAGCTCCCTCCTACTT 7346

```

RESULT 15

AL391843/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

```

AL391843 47972 bp DNA PRI 22-FEB-2001
Human DNA sequence from clone RP11-592B10 on chromosome 6, complete
sequence.
AL391843
AL391843.13 GI:13157590
HTG.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 47972)
Sycamore, N.
Direct Submission
Submitted (22-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonequests@sanger.ac.uk
On Feb 27, 2001 this sequence version replaced GI:12831877.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with

```

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWSPPROT; Tr: TREMBL; Wp: WORMPEP; information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr6>

RP11-592B10 is from the library RP11-11.3 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/dacpac/home.htm>

VECTOR: pBAC3.6

FEATURES

source

1. 47972

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/clone="RP11-592B10"

/clone_lib="RP11-11.3"

/complement(147..519)

/note="match: GSS: Em:AQ253031"

156..328

/note="match: GSS: Em:AQ542267"

complement(136..518)

/note="match: GSS: Em:AQ210521"

complement(156..512)

/note="match: GSS: Em:AQ560490"

164..366

/note="LTR12 repeat: matches 18..192 of consensus"

332..503

/note="match: GSS: Em:AQ542267"

427..472

/note="LTR12 repeat: matches 189..234 of consensus"

473..549

/note="LTR30 repeat: matches 544..620 of consensus"

480..583

/note="MER61E repeat: matches 377..477 of consensus"

584..673

/note="LTR12 repeat: matches 582..671 of consensus"

674..7342

/note="HERV22 repeat: matches 100..6836 of consensus"

7344..7825

/note="LTR22 repeat: matches 1..509 of consensus"

7826..11300

/note="LIM1 repeat: matches 685..4254 of consensus"

10015..10095

/note="single clone region. Assembly confirmed by restriction digest data."

11301..11611

/note="AluY repeat: matches 1..309 of consensus"

11612..13713

/note="LIM1 repeat: matches 4254..6304 of consensus"

13766..15206

/note="LIM3 repeat: matches 2143..3566 of consensus"

15207..15512

/note="AluX repeat: matches 2..308 of consensus"

15513..17628

/note="LIM3 repeat: matches 3546..5738 of consensus"

18106..18525

/note="Tiger2a repeat: matches 23..433 of consensus"

19408..20039

/note="LTR9 repeat: matches 6..625 of consensus"

20213..20397

/note="LIM4 repeat: matches 4454..4646 of consensus"

20556..21581

/note="MER41B repeat: matches 1..635 of consensus"

22205..22776

/note="MER77 repeat: matches 13..629 of consensus"

23549..24036

/note="match: GSS: Em:AQ66040"

23847..23903

/note="L2 repeat: matches 2639..2695 of consensus"

complement(24422..24864)

/note="match: GSS: Em:AQ610704"

complement(24640..24857)

/note="match: GSS: Em:AQ030189"

24874..24928

/note="L2 repeat: matches 2692..2746 of consensus"

complement(25273..25508)

/note="match: GSS: Em:AQ376698"

complement(25257..25473)

/note="match: GSS: Em:AQ545471"

complement(25299..25496)

/note="match: GSS: Em:AQ749890"

complement(25327..25490)

/note="match: GSS: Em:B56335"

26210..26253

/note="22 copies 2 mer tg 95% conserved"

26543..26868

/note="match: GSS: Em:AQ902789"

26950..27010

/note="MER90 repeat: matches 3..63 of consensus"

27054..27374

/note="MER90 repeat: matches 222..543 of consensus"

27375..27869

/note="LTR15 repeat: matches 1..489 of consensus"

27870..27942

/note="MER90 repeat: matches 543..615 of consensus"

28722..28935

/note="AluY repeat: matches 74..286 of consensus"

complement(28723..29206)

/note="match: GSS: Em:AQ82343"

30251..30554

/note="AluY repeat: matches 1..305 of consensus"

33533..33906

/note="LTR12 repeat: matches 1..374 of consensus"

34190..34587

/note="LTR12 repeat: matches 117..513 of consensus"

34664..35066

/note="MSTR repeat: matches 1..424 of consensus"

35260..35321

/note="31 copies 2 mer aa 80% conserved"

35346..35777

/note="LTR1 repeat: matches 1..437 of consensus"

35952..36169

/note="LTR1A repeat: matches 1..210 of consensus"

36201..36250

/note="25 copies 2 mer ca 82% conserved"

36271..36393

/note="FLAN_A repeat: matches 2..124 of consensus"

37540..37613

/note="37 copies 2 mer aa 67% conserved"

38348..38399

/note="26 copies 2 mer at 84% conserved"

complement(39249..40015)

/note="match: GSS: Em:AQ38237"

40021..40482

/note="match: GSS: Em:AQ171081"

40101..40352

/note="WIR repeat: matches 1..259 of consensus"

40581..40676

/note="48 copies 2 mer aa 64% conserved"

```

repeat_region 41234..41662
/note="LJM2 repeat: matches -676..-231 of consensus"
repeat_region 42139..42439
/note="LIPAL5-16 repeat: matches -265..-24 of consensus"
repeat_region 42901..43464
/note="LIPAL3 repeat: matches 802..1288 of consensus"
repeat_region 43465..43687
/note="MR30 repeat: matches 1..230 of consensus"
repeat_region 43688..43754
/note="LIPAL3 repeat: matches 1288..1351 of consensus"
repeat_region 43755..46738
/note="LIPAL6 repeat: matches 3178..6153 of consensus"
repeat_region 46921..47311
/note="LIPAL6 repeat: matches 1..304 of consensus"
misc_feature 47394
/note="match: GSS: Em:A053386"
complement(47357..47846)
/note="match: GSS: Em:A0534016"
47878..47972
/note="match: GSS: Em:A0186739"
47880..47972
/note="match: GSS: Em:A0551896"
BASE COUNT 16896 a 8982 c 8503 g 13591 t
ORIGIN

```

```

Query Match 84.8% Score 17.8; DB 90; Length 47972;
Best Local Similarity 90.5%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 aatgaactccctgtactctt 21
|||||
Db 29145 AAGTGAACCTCCTCCTT 29125

```

Search completed: July 25, 2001, 05:16:42
Job time: 9205 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 04:05:07 ; Search time 247.87 Seconds

(without alignments)
53.197 Million cell updates/sec

Title: US-09-142-095-1

Perfect score: 21
Sequence: 1 aagtaactccctgcactctt 21

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 7301 segs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	AAAT79541	UGT1*1 gene exon 1
2	21	100.0	21	AAAT79541	Upstream DNA seque
3	16.8	80.0	237326	AAV57903	Hereditary haemoch
4	16.2	77.1	260	AAAC25716	Human secreted pro
5	16.2	77.1	867	AAAC38612	Arabidopsis thalia
6	16.2	77.1	1613	AAAC45500	CDNA encoding enzy
7	16.2	77.1	32042	AAZ09252	Human CARD-4 gene.
8	16.2	77.1	32042	AAZ09252	Human CARD-4 gene.
9	15.8	75.2	957	AAO97865	TRONP1 DNA, TRepo
10	15.8	75.2	957	AAO97865	Treponea pallidum
11	15.2	72.4	537	AAAG7251	Eucalyptus grandis

12	15.2	72.4	640	AAV04426	Human calpain smal
13	15.2	72.4	804	AAV04420	Human calpain smal
14	15.2	72.4	1242	AACT7765	Human cancer assoc
15	15.2	72.4	1741	AAZ98118	Human signal pepti
16	15.2	72.4	1938	AAAC6144	Arabidopsis thalia
17	15.2	72.4	1940	AAAC39039	Arabidopsis thalia
18	15.2	72.4	2024	AAAC81718	Human secreted pro
19	15.2	72.4	2582	AAV99562	Soybean lysine ket
20	15.2	72.4	4044	AAAC81899	A. thaliana Srp30
21	15.2	72.4	4665	AAV97955	Human epididymis-s
22	15.2	72.4	4665	AAV97955	Hereditary haemoch
23	15.2	72.4	235033	AAV57926	Arabidopsis thalia
24	15.2	72.4	1330	AAAC33777	Polynucleotide seg
25	14.8	70.5	307	AAAC8243	Human colon cancer
26	14.8	70.5	413	AAAC26942	Human secreted pro
27	14.8	70.5	731	AAAT20482	Human gene signatu
28	14.8	70.5	1398	AAAX25083	Potato tuber-speci
29	14.8	70.5	1723	AAAT42133	Ascorbate-free-rad
30	14.8	70.5	2161	AAZ30444	Tobacco anthranila
31	14.8	70.5	2526	AAZ20026	Human rad17 cell c
32	14.8	70.5	2535	AAZ20033	Human rad17 cell c
33	14.8	70.5	2540	AAZ20032	Human rad17 cell c
34	14.8	70.5	2543	AAZ20051	Human rad17 cell c
35	14.8	70.5	2647	AAZ20035	Human rad17 cell c
36	14.8	70.5	2652	AAZ44655	Testis-associated
37	14.8	70.5	2977	AAAT30085	Human Raf1 kinase
38	14.8	70.5	2977	AAAT61894	Human Raf-1 CDNA.
39	14.8	70.5	2977	AAV20439	Human c-raf oncoge
40	14.8	70.5	2977	AAAX9340	Human c-raf sequen
41	14.8	70.5	2977	AAAT8137	Human c-raf-1 gene
42	14.8	70.5	2977	AAAT3552	Polynucleotide #1
43	14.8	70.5	2977	AAAT8137	Human c-raf DNA.
44	14.8	70.5	2977	AAAT8137	Human c-raf CDNA.
45	14.8	70.5	5907	AACT6910	Human OREX ORF2465

ALIGNMENTS

RESULT 1
AAAT79541 standard: DNA; 21 BP.
ID AAAT79541
AC AAAT79541:
XX 23-JAN-1998 (first entry)
DT UGT1*1 gene exon 1 upstream PCR primer A.
DE XX
KM Uridine diphosphate glucuronosyltransferase gene; UGT1
KM Gilbert's syndrome; GS; unconjugated hyperbilirubinaemia;
KM bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;
KM Drug trial efficiency; screening; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
XX WO9732042-A2.
XX
XX 04-SEP-1997.
XX
XX 03-MAR-1997; 97NO-GB00577.
XX
XX 16-MAR-1996; 96GB-0005598.
XX
XX 01-MAR-1996; 96GB-0004480.
XX
XX (UYDU-) UNIV DUNDEE.
XX
XX Burchell B;
XX
XX WPI; 1997-448702/41.
XX
XX Improving drug trial efficiency comprises identifying participants

PT with Gilbert's syndrome - useful as their altered drug metabolism
 PT may hinder result interpretation
 XX Claim 14; Page 12; 31pp; English.
 XX This PCR primer (with primers AAT79542-44) flanks the TATA box sequence
 CC upstream of the uridine diphosphate glucuronosyltransferase (UGT) gene
 CC 1*1 exon 1 (see AAT79540), and was used to amplify fragments of 253-255
 CC bp. This gene is known to be associated with Gilbert's syndrome (GS). GS
 CC is a mild, common form of unconjugated hyperbilirubinaemia associated
 CC with reduced bilirubin glucuronidation capacity. Analysis of the genetic
 CC basis of GS has allowed 2 forms to be identified. One is a mild form
 CC associated with a homozygous 2 bp insertion in the TATA sequence
 CC upstream of the UGT*1 exon 1, and the other is a more severe form
 CC associated with heterozygosity for a mutation which, when homozygous,
 CC causes Crigler-Najjar type 2 disease. The first form is autosomal
 CC recessive and the second is inherited dominantly. Patients suffering from
 CC GS, which is benign, may have altered metabolism of some drugs, making it
 CC difficult to determine if an effect is due to the drug or the syndrome.
 CC Drug trial efficiency would be improved if potential participants can be
 CC screened for the genetic basis of GS, and eliminated or included on
 CC basis of them possessing or not possessing GS.
 CC
 CC Sequence 21 BP; 5 A; 7 C; 3 G; 6 T; 0 other;
 SQ
 Query Match 100.0%; Score 21; DB 18; Length 21;
 Best Local Similarity 100.0%; Pred. No. 0.14; 0; Indels 0; Gaps 0;
 Matches 21; Conservative 0; Mismatches 0;
 OY 1 aagtgacctcctgctactct 21
 DB 1 aagtgacctcctgctactct 21
 RESULT 2
 AAT79540
 ID AAT79540 standard; DNA; 620 BP.
 XX
 XX AAT79540:
 AC
 XX 23-JAN-1998 (first entry)
 DT
 XX
 XX Upstream DNA sequence of UGT*1 gene exon 1.
 DE
 XX
 XX uridine diphosphate glucuronosyltransferase gene; UGT;
 KW Gilbert's syndrome; GS; unconjugated hyperbilirubinaemia;
 KW bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;
 KW Drug trial efficiency; screening; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT protein_bind 2..8
 FT /tag- a
 FT /bound_moely- AP1
 FT GC_signal 93..102
 FT /tag- b
 FT /note- "SP1 binding site"
 FT misc-feature 180..191
 FT /tag- c
 FT /note- "feature indicated in patent, but no further
 FT explanation is given"
 FT
 FT protein_bind 315..323
 FT /tag- d
 FT /bound_moely- AP3
 FT protein_bind 317..324
 FT /tag- e
 FT /bound_moely- CLBP
 FT misc-feature 366..376
 FT /tag- f
 FT /note- "feature indicated in patent, but no further
 FT explanation is given"
 FT

FT protein_bind 386..392
 FT /tag- g
 FT /bound_moely- AP1
 FT protein_bind 513..519
 FT /tag- h
 FT /bound_moely- AP1
 FT misc-feature 513..520
 FT /tag- i
 FT /note- "feature indicated in patent, but no further
 FT explanation is given"
 FT
 FT protein_bind 520..532
 FT /tag- j
 FT /bound_moely- HNF1
 FT TATA_signal 558..572
 FT /tag- k
 FT /note- "corresponds to positions -53 to -39 in patent"
 FT
 FT misc-feature 595
 FT /tag- l
 FT /note- "feature indicated in patent, but no further
 FT explanation is given"
 FT
 FT CDS 612
 FT /tag- m
 FT /codon_start- 612
 FT
 FT WO9732042-A2.
 PN
 XX
 XX 04-SEP-1997.
 PD
 XX 03-MAR-1997; 97WO-GB00577.
 PE
 XX 16-MAR-1996; 96GB-0005598.
 PR 01-MAR-1996; 96GB-0004480.
 XX
 XX (UYDU-) UNIV DUNDEE.
 PA
 XX
 XX Burchell B;
 PI
 XX
 XX WPI: 1997-448702/41.
 DR
 XX
 XX
 XX Improving drug trial efficiency comprises identifying participants
 PT with Gilbert's syndrome - useful as their altered drug metabolism
 PT may hinder result interpretation
 XX
 XX Claim 11; Fig 4; 31pp; English.
 PS
 XX
 XX This sequence represents the upstream sequence, positions -611 to 9
 CC in the patent, of uridine diphosphate glucuronosyltransferase (UGT)
 CC gene 1*1 exon 1. This gene is known to be associated with Gilbert's
 CC syndrome (GS). GS is a mild, common form of unconjugated
 CC hyperbilirubinaemia associated with reduced bilirubin glucuronidation
 CC capacity. Analysis of the genetic basis of GS has allowed 2 forms to be
 CC identified. One is a mild form associated with a homozygous 2 bp
 CC insertion in the TATA sequence upstream of the UGT*1 exon 1, and
 CC the other is a more severe form associated with heterozygosity for a
 CC mutation which, when homozygous, causes Crigler-Najjar type 2 disease.
 CC The first form is autosomal recessive and the second is inherited
 CC dominantly. Patients suffering from GS, which is benign, may have
 CC altered metabolism of some drugs, making it difficult to determine if an
 CC effect is due to the drug or the syndrome. Drug trial efficiency would
 CC be improved if potential participants can be screened for the genetic
 CC basis of GS, and eliminated or included on basis of them possessing or
 CC not possessing GS. In this case, screening involves PCR amplification
 CC of the UGT gene, using the primers described in AAT79541-44.
 CC
 XX
 XX Sequence 620 BP; 157 A; 127 C; 151 G; 185 T; 0 other;
 SQ
 Query Match 100.0%; Score 21; DB 18; Length 620;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 aagtgacctcctgctactct 21
 |||||||||||||||||||||

DB 465 aagtaactccctgctacctt 465

RESULT 3

AAV57903
ID AAV57903 standard; DNA: 237326 BP.

AAV57903;

21-DEC-1998 (first entry)

Hereditary haemochromatosis subregion from an HH affected individual.

Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;

diagnosis; iron metabolism; NPT3; NPT4; Roret; BTF1; BTF2; BTF3;

BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;

type 1 sodium transport gene; ss.

Homo sapiens.

WO9814466-A1.

09-APR-1998.

30-SEP-1997; 97MO-US17658.

07-MAY-1997; 97US-0852495.

01-OCT-1996; 96US-0724394.

(PROG-) PROGENITOR INC.

Feder JN, Krommal GS, Lauer PM, Ruddy DA, Thomas WT;

Tsuchihashi Z, Wolff RK;

WPI, 1998-240014/21.

Hereditary haemochromatosis gene products - used to develop products

for the diagnosis and treatment of hereditary disorders in iron

metabolism

Claim 1; Fig 9; 209pp; English.

The present invention describes hereditary haemochromatosis gene

products from the human haemochromatosis gene. The present sequence

represents a hereditary haemochromatosis subregion from an hereditary

haemochromatosis (HH) affected individual. Also described is a

method to determine the presence or absence of the common hereditary

haemochromatosis (HFE) gene mutation in an individual comprising:

(a) providing DNA or RNA from the individual; and (b) assessing the

DNA or RNA for the presence or absence of a haplotype or genotype where

the presence or absence of the haplotype genotype indicates the likely

presence of the HFE gene mutation in the genome of the individual. The

HFE gene sequences from the present invention can be used to develop

products for use in the diagnosis and treatment of HFE. The present

invention also describes BTF genes, which are homologues of the milk

protein butyrophilin (BTF), and can be used in the production of agonists

and antagonists of BTF function. Also described are: (1) a Roret gene

which can be used to develop products for the study, diagnosis and

treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes

which are homologues of a type 1 sodium transport gene, and can

similarly be used for hypophosphatemia.

Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T; 0 other:

Query Match 80.0%; Score 16.8; DB 19; Length 237326;

Best Local Similarity 90.0%; Pred. No. 53;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 aagtaactccctgctacctt 20

DB 105327 aagtaactccctgctacctt 105346

RESULT 4

AAC25716/c

ID AAC25716 standard; cDNA: 260 BP.

AAC25716;

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 29791.

Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST) GENSET.

Dumas Malne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

Claim 1; SEQ ID 29791; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from

mRNAs encoding secreted proteins. No ORF has yet been conclusively

identified within the present sequence. The 5' ESTs were prepared from

total human RNAs or poly(A) RNAs derived from 30 different tissues. EST

sequences usually correspond mainly to the 3' untranslated region (UTR)

of the mRNA because they are often obtained from oligo-dT primed cDNA

libraries. Such ESTs are not well suited for isolating cDNA sequences

derived from the 5' ends of mRNAs and even in those cases where longer

cDNA sequences have been obtained, the full 5' UTR is rarely included.

5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

in diagnostic, forensic, gene therapy and chromosome mapping procedures.

They are used to obtain upstream regulatory sequences and to design

expression and secretion vectors.

Sequence 260 BP; 74 A; 58 C; 45 G; 82 T; 1 other:

Query Match 77.1%; Score 16.2; DB 21; Length 260;

Best Local Similarity 85.7%; Pred. No. 46;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 aagtaactccctgctacctt 21

DB 68 AAGGAACCTGCTCTACCTT 48

RESULT 5

AAC38612/c

ID AAC38612 standard; DNA: 867 BP.

AAC38612;

17-OCT-2000 (first entry)

Arabidopsis thaliana DNA fragment SEQ ID NO: 21606.

Hybridization assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway;
metabolic pathway; promoter; termination sequence; ss.
Arabidopsis thaliana.
EPI033405-A2.
06-SEP-2000.
25-FEB-2000; 2000EP-0301439.
25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126264.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
15-APR-1999; 99US-0129845.
21-APR-1999; 99US-0130077.
23-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
28-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
30-APR-1999; 99US-0132407.
04-MAY-1999; 99US-0132484.
05-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134768.
18-MAY-1999; 99US-0134769.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135553.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140354.
23-JUN-1999; 99US-0140695.
24-JUN-1999; 99US-0140823.
28-JUN-1999; 99US-0140991.
29-JUN-1999; 99US-0141287.
30-JUN-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147038.
04-AUG-1999; 99US-0147204.
04-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147418.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148341.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148684.
16-AUG-1999; 99US-0149366.
17-AUG-1999; 99US-0149175.
18-AUG-1999; 99US-0149426.
20-AUG-1999; 99US-0149722.
20-AUG-1999; 99US-0149723.
20-AUG-1999; 99US-0149920.
23-AUG-1999; 99US-0149902.
23-AUG-1999; 99US-0149930.
25-AUG-1999; 99US-0150566.
26-AUG-1999; 99US-0150884.
27-AUG-1999; 99US-0151065.
27-AUG-1999; 99US-0151066.
27-AUG-1999; 99US-0151080.
30-AUG-1999; 99US-0151303.
31-AUG-1999; 99US-0151438.

PR	01-SEP-1999;	99US-0151930;
PR	07-SEP-1999;	99US-0152363;
PR	10-SEP-1999;	99US-0153070;
PR	13-SEP-1999;	99US-0153758;
PR	15-SEP-1999;	99US-0154018;
PR	16-SEP-1999;	99US-0154039;
PR	20-SEP-1999;	99US-0154779;
PR	22-SEP-1999;	99US-0155139;
PR	23-SEP-1999;	99US-0155486;
PR	24-SEP-1999;	99US-0155659;
PR	28-SEP-1999;	99US-0156458;
PR	29-SEP-1999;	99US-0156596;
PR	04-OCT-1999;	99US-0157117;
PR	05-OCT-1999;	99US-0157753;
PR	06-OCT-1999;	99US-0157863;
PR	07-OCT-1999;	99US-0158029;
PR	08-OCT-1999;	99US-0158332;
PR	12-OCT-1999;	99US-0158369;
PR	13-OCT-1999;	99US-0159293;
PR	13-OCT-1999;	99US-0159294;
PR	13-OCT-1999;	99US-0159325;
PR	14-OCT-1999;	99US-0159329;
PR	14-OCT-1999;	99US-0159330;
PR	14-OCT-1999;	99US-0159331;
PR	14-OCT-1999;	99US-0159637;
PR	14-OCT-1999;	99US-0159638;
PR	18-OCT-1999;	99US-0159584;
PR	21-OCT-1999;	99US-0160741;
PR	21-OCT-1999;	99US-0160767;
PR	21-OCT-1999;	99US-0160768;
PR	21-OCT-1999;	99US-0160770;
PR	21-OCT-1999;	99US-0160814;
PR	21-OCT-1999;	99US-0160815;
PR	22-OCT-1999;	99US-0160825;
PR	22-OCT-1999;	99US-0160980;
PR	22-OCT-1999;	99US-0160981;
PR	22-OCT-1999;	99US-0160989;
PR	25-OCT-1999;	99US-0161404;
PR	25-OCT-1999;	99US-0161405;
PR	25-OCT-1999;	99US-0161406;
PR	26-OCT-1999;	99US-0161359;
PR	26-OCT-1999;	99US-0161360;
PR	26-OCT-1999;	99US-0161361;
PR	28-OCT-1999;	99US-0161920;
PR	28-OCT-1999;	99US-0161992;
PR	28-OCT-1999;	99US-0161993;
PR	29-OCT-1999;	99US-0162142;

```

Query Match          77.1%; Score 16.2; DB 21; Length 867;
Best Local Similarity 85.7%; Pred. NO. 53;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 aaagtgaactccctgactacctt 21
        ||| ||||| ||| ||| |||
Db       678 AAGTGAACCTCCTGCGACTTT 658

RESULT    6
AAA46500/C
ID   AAA46500 standard; cDNA; 1613 BP.
XX
XX
AA46500;
AC
CC
XX
DT   04-SEP-2000 (first entry)
DE
XX
CDNA encoding enzyme involved in mannan polysaccharide hydrolysis.
XX
KW   Hydrolysis; polysaccharide; mannan; coffee; endo-beta-mannanase; ds.
XX
OS   Coffea arabica.
XX
Key      Location/Qualifiers
FT       11..1294
            /tag= a

```

/product- "endo-beta-mannanase"
 WO200028046-A1.
 18-MAY-2000.
 28-OCT-1999; 99WO-EP08314.
 11-NOV-1998; 98EP-0203742.
 (NEST) SOC PROD NESTLE SA.
 Marraccini P, Rogers J;
 WPI: 2000-399535/34.
 P-PSDB; AMY93441.
 New DNA encoding endo-beta-mannanase from coffee, used e.g. in
 pharmaceutical, cosmetic or food compositions to hydrolyze polymanans
 .
 Claim 4; Page 29-30; 41pp; French.
 The present sequence encodes an endo-beta-mannanase enzyme, which is
 involved in the hydrolysis of polysaccharides that consist of molecules
 of mannan, either simple or branched, linked together by beta(1-4) bonds.
 The mannanase polynucleotide sequence is used for in vivo modification
 of the coffee endo-beta-mannanase gene. It is also used to produce
 transgenic plant cells (especially coffee cells) which have modified
 properties of mannan polysaccharide, and thus altered flavour or
 structure. The enzyme is used for modification, degradation or synthesis
 of mannan polysaccharides in vitro, particularly to treat coffee beans
 to increase the percentage of dry matter extraction, and thus reduce the
 quantity of sediment.
 Sequence 1613 BP; 520 A; 305 C; 347 G; 441 T; 0 other:

Query Match	77.1%	Score 16.2	DB 21	Length 1613
Best Local Similarity	85.7%	Pred. No. 58		
Matches 18	Conservative 0	Mismatches 3	Indels 0	Gaps 0
<p>Oy 1 aagtaacctccgcctacctt 21 </p>				
Db 408 AAGTGAACCTCCACTTCCTT 308				
RESULT 7				
AAZ09252				
ID AAZ09252 standard; DNA; 32042 BP.				
AAZ09252;				
25-OCT-1999 (first entry)				
XX Human CARD-4 DNA.				
XX CARD-3; caspase recruitment domain; CARD-4; regulation; detection;				
XX caspase activation; detection; screening; therapy; diagnosis; disease;				
XX apoptotic cell death; Fas/APD-1 receptor complex; TNF receptor complex;				
XX cancer; follicular lymphoma; carcinoma; p53 mutation; viral infection;				
XX hormone-dependent tumour; autoimmune disorder; Alzheimer's disease;				
XX systemic lupus erythematosus; immune-mediated glomerulonephritis; stroke;				
XX Parkinson's disease; amyotrophic lateral sclerosis; retinitis pigmentosa;				
XX spinal muscular dystrophy; cerebellar degeneration; anaemia; drug;				
XX myelodysplastic syndrome; myocardial infarction; cell proliferation;				
XX cell differentiation; cell survival; CARD-4L; CARD-4S; CARD-4Y;				
XX CARD-4Z; human; ds.				
XX Homo sapiens.				
XX OS				
XX Key Location/Qualifiers				
XX exon 364..685				

PR	06-FEB-1998;	98US-001994Z.
PR	17-JUN-1998;	98US-009904I.
PA	(MILL-) MILLENNium PHARM INC.	
PX	Bertin J;	
DR	WPI; 1999-49A269/41.	
XX		
PT	Novel CARD-3 and CARD-4 genes and polypeptides used or treating regulation of cellular proliferation and differentiation and cell survival	
PS	Example 13; Fig 18; 181pp; English.	
XX		
CC	This invention describes the isolation of novel human caspase recruitment domain, CARD-3 and CARD-4 polynucleotides and proteins and a partial murine CARD-4 protein and genes. The genes and proteins of the invention are involved in the regulation of caspase activation. The caspase recruitment domain (CARD) polynucleotides, polypeptides, homologues and antibodies can be used in screening assays, detection assays, predictive medicine and therapeutic and prophylactic methods of treatment. The methods may be used to diagnose and treat patients which are suffering from a disorder associated with abnormal level or rate of apoptotic cell death, abnormal activity of the Fas/Apo-1 receptor complex, abnormal activity of the TNF receptor complex, or abnormal activity of a caspase. Diseases that may be treated include cancer (particularly follicular lymphoma, carcinomas associated with mutations in p53 and hormone-dependent tumours), autoimmune disorders (e.g. systemic lupus erythematosus, immune-mediated glomerulonephritis). Viral infections, Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, retinitis pigmentosa, spinal muscular dystrophy, cerebellar degeneration, anaemia, myelodysplastic syndrome, myocardial infarction, and stroke. CARD-3 protein interacts with other cellular proteins, and so can be used for regulation of cellular proliferation and differentiation and cell survival. The CARD proteins may also be used to for screen drugs or compounds which modulate their activity. The CARD-4 gene can express a long transcript that encodes CARD-4L, a short transcript that encodes CARD-4S or two CARD-4 splice variants, CARD-4Y and CARD-4Z. This sequence represents a genomic DNA sequence which encodes the human CARD-4 protein.	
SQ	Sequence 32042 BP; 7389 A; 7540 C; 7721 G; 9392 T; 0 other;	
Query Match	77.1%; Score 16.2; DB 20; Length 32042;	
Best Local Similarity	85.7%; Pseq. No. 83;	
Matches 18; Conservative 0; Mismatches 3; Indels 0; Caps 0;		
OY	1 aaagtaactcctgccttcctt 21 	
Db	17935 aagggaactccctgatacctt 17955	
AALRESULT 8		
ID	AAF30011 standard; cDNA; 32042 BP.	
XX		
AC	AAF30011;	
XX		
DT	23-APR-2001 (first entry)	
DE		
XX		
XX	Human CARD-4 gene.	
KW	CARD-4; caspase recruitment domain; human; cancer; infection; autoimmune disease; neurological disease; haematological disease; immune disease; inflammation; antitumour; antiapoptic; immunomodulator; antiinflammatory; apoptosis; diagnosis; gene therapy; chromosome 7; ds.	
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	485..31768

```

FT      /*tag- a
FT      /note- "Contains Introns"
FT      exon
FT      /*tag- b
FT      /number- 1
FT      686..2094
FT      Intron
FT      /*tag- c
FT      /number- 1
FT      2095..2269
FT      exon
FT      /*tag- d
FT      /number- 2
FT      2270..4365
FT      Intron
FT      /*tag- e
FT      /number- 2
FT      4366..6190
FT      exon
FT      /*tag- f
FT      /number- 3
FT      6191..9024
FT      Intron
FT      /*tag- g
FT      /number- 3
FT      9025..9108
FT      exon
FT      /*tag- h
FT      /number- 4
FT      9109..10355
FT      Intron
FT      /*tag- i
FT      /number- 4
FT      10356..10439
FT      exon
FT      /*tag- j
FT      /number- 5
FT      10440..11181
FT      Intron
FT      /*tag- k
FT      /number- 5
FT      11182..11265
FT      exon
FT      /*tag- l
FT      /number- 6
FT      11266..19749
FT      Intron
FT      /*tag- m
FT      /number- 6
FT      19750..19833
FT      exon
FT      /*tag- n
FT      /number- 7
FT      19834..21324
FT      Intron
FT      /*tag- o
FT      /number- 7
FT      21325..21408
FT      exon
FT      /*tag- p
FT      /number- 8
FT      21409..24226
FT      Intron
FT      /*tag- q
FT      /number- 8
FT      24227..24310
FT      exon
FT      /*tag- r
FT      /number- 9
FT      24311..27948
FT      Intron
FT      /*tag- s
FT      /number- 9
FT      27949..28032
FT      exon
FT      /*tag- t
FT      /number- 10
FT      28033..31695
FT      Intron
FT      /*tag- u
FT      /number- 10
FT      31696..32024
FT      exon
FT      /*tag- v
FT      /number- 11

```

```

XX      /*tag- a
XX      /note- "Contains Introns"
XX      exon
XX      /*tag- b
XX      /number- 1
XX      686..2094
XX      Intron
XX      /*tag- c
XX      /number- 1
XX      2095..2269
XX      exon
XX      /*tag- d
XX      /number- 2
XX      2270..4365
XX      Intron
XX      /*tag- e
XX      /number- 2
XX      4366..6190
XX      exon
XX      /*tag- f
XX      /number- 3
XX      6191..9024
XX      Intron
XX      /*tag- g
XX      /number- 3
XX      9025..9108
XX      exon
XX      /*tag- h
XX      /number- 4
XX      9109..10355
XX      Intron
XX      /*tag- i
XX      /number- 4
XX      10356..10439
XX      exon
XX      /*tag- j
XX      /number- 5
XX      10440..11181
XX      Intron
XX      /*tag- k
XX      /number- 5
XX      11182..11265
XX      exon
XX      /*tag- l
XX      /number- 6
XX      11266..19749
XX      Intron
XX      /*tag- m
XX      /number- 6
XX      19750..19833
XX      exon
XX      /*tag- n
XX      /number- 7
XX      19834..21324
XX      Intron
XX      /*tag- o
XX      /number- 7
XX      21325..21408
XX      exon
XX      /*tag- p
XX      /number- 8
XX      21409..24226
XX      Intron
XX      /*tag- q
XX      /number- 8
XX      24227..24310
XX      exon
XX      /*tag- r
XX      /number- 9
XX      24311..27948
XX      Intron
XX      /*tag- s
XX      /number- 9
XX      27949..28032
XX      exon
XX      /*tag- t
XX      /number- 10
XX      28033..31695
XX      Intron
XX      /*tag- u
XX      /number- 10
XX      31696..32024
XX      exon
XX      /*tag- v
XX      /number- 11

```

```

XX      (MILL-) MILLENNIUM PHARM INC.
XX      BERTIN J.
XX      WPI: 2001-061973/07.
XX      P-PSDB: AAB20080, AAB20081, AAB20082, AAB20083.
XX      Isolated intracellular proteins predicted to be involved in regulating
XX      caspase activation are used for diagnosis and treatment of e.g. cancer,
XX      viral infections, autoimmune diseases, neurological diseases and
XX      haematological disorders.
XX      Example 13: Fig 18; 208pp; English.
XX      The present sequence is that of the human caspase recruitment
XX      domain 4 (CARD-4) gene on chromosome 7. CARD-4 exists in at least
XX      4 forms, i.e. the long form CARD-4L (see AAB20080), the short form
XX      CARD-4S (see AAB20081), and splice variants CARD-4Y (see AAB20082)
XX      and CARD-4Z (see AAB20082). It is an intracellular protein
XX      predicted to be involved in regulating caspase activation. It
XX      activates the NF-kappa pathway and enhances caspase-9-mediated
XX      cell death. Methods of diagnosing and treating patients suffering
XX      from a disorder associated with an abnormal level or rate of apoptotic
XX      cell death, abnormal activity of the Fas/Apo-1 receptor complex,
XX      abnormal activity of the tumour necrosis factor receptor complex,
XX      or abnormal activity of a caspase involve administering a compound
XX      that modulates the expression or activity of CARD-3, CARD-4, CARD-5
XX      or CARD-6 e.g. a small molecule, antisense nucleic acid, ribozyme
XX      or polypeptide. Such disorders include cancer, viral infection,
XX      autoimmune disorders, neurological diseases, haematological
XX      disorders, inflammatory disorders and immune disorders. The CARD-4
XX      gene is useful for genetic information and mapping and identifying
XX      mutations, e.g. mutations in splice donor or acceptor sites.
XX      Sequence 32042 BP; 7389 A; 7540 C; 7721 G; 9392 T; 0 other;
XX
XX      Query Match          77.1%; Score 16.2; DB 22; Length 32042;
XX      Best Local Similarity 85.7%; Pred. No. 83;
XX      Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX      OY      1 aagtgaaacctccgtgacctt 21
XX      Db      17935 aagsgaaacctccgtgacctt 17955
XX
XX      RESULT 9
XX      AA097865
XX      ID      AA097865 standard; DNA; 957 BP.
XX      AC      AA097865;
XX      DT      05-DEC-1995 (first entry)
XX      DE      TROMP1 DNA.
XX      OS      Treponema pallidum subsp. pallidum.
XX      KW      TROMP1; rare outer membrane protein; Spirochaetaceae; immunogen;
XX      vaccine; syphilis; ss.
XX      Key      Location/Qualifiers
XX      CDS      1..957
XX      sig_peptide 1..96
XX      mat_peptide 97..954
XX      FT      /*tag- a
XX      FT      /*tag- b
XX      FT      /*tag- c
XX      PD      M03518633-A.
XX      13-JUL-1995.

```

XX 06-JAN-1995; 95WO-US00190.
 PF 17-AUG-1994; 94US-0292904.
 XX 06-JAN-1994; 94US-0178084.
 PR 07-JUN-1994; 94US-0255322.
 XX (REGC) UNIV CALIFORNIA.
 PA Blanco DR, Champion CT, Lovett MA, Miller JN;
 PI WPI; 1995-254907/33.
 DR P-PSDB; AAR79722.
 XX Rare outer coat membranes of Spirochaetaceae prodn. - by density
 PT gradient fractionation, also isolated immunogenic proteins for use in
 PT vaccines, derived antibodies and nucleic acid.
 PS Disclosure; Page 48-50; 78pp; English.
 CC TROMP1 (given in AAR79722) is encoded by DNA isolated using tryptic
 CC digest amino acid sequence analysis of EcoRI-digested genomic DNA
 CC of *T. pallidum* subsp. *pallidum*. It is the precursor of a 31 kDa rare
 CC outer membrane protein useful for prepn. of vaccine against syphilis
 CC and other treponemal diseases.
 XX Sequence 957 BP; 210 A; 206 C; 303 G; 238 T; 0 other;
 SQ

Query Match 75.2%; Score 15.8; DB 16; Length 957;
 Best Local Similarity 89.5%; Pred. No. 85;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 3 gtaactccctgctactct 21
 |||||||||||||
 Db 749 gtaactccctgctactct 767

RESULT 10
 AAT6580
 ID AAT6580 standard; DNA; 957 BP.
 XX AAT6580:
 AC 13-JAN-1998 (first entry)
 XX
 DT Treponema pallidum rare outer membrane protein (TROMP-1) DNA.
 DE
 XX
 XX Rare; outer membrane protein; Treponema pallidum; Spirochaetales;
 KM Immune response; Syphilis; TROMP-1; precursor; ss.
 XX
 OS Treponema pallidum.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..957
 FT /tag- a
 FT /product- Immature_TROMP1_protein
 FT
 XX MO9727225-A1.
 PN
 XX 31-JUL-1997.
 PD
 XX 23-JAN-1997; 97WO-US01302.
 PF
 XX 23-JAN-1996; 96US-0599480.
 PR
 XX (SLOK) SIOAN KETTERING INST CANCER RES.
 PA (REGC) UNIV CALIFORNIA.
 PI Blanco DR, Lovitt MA, Miller JN, Tempst PJ;
 DR WPI; 1997-393614/36.
 DR P-PSDB; AAR22134.

XX New isolated Treponema pallidum outer membrane protein - used in the
 PT detection and induction of immune response in an animal to
 PT pathogenic Spirochaetales
 XX Disclosure; Pages 51-53; 77pp; English.
 PS
 XX This 957 bp sequence (TROMP1 gene) encodes a precursor rare outer
 CC membrane protein (OMP) of species Treponema pallidum, subspecies
 CC pallidum. The TROMP1 gene was cloned in a procedure where mixed
 CC oligonucleotides (31-A,31-C) hybridized to a EcoRI restriction
 CC fragment by Southern blot analysis of *T. pallidum* genomic DNA. DNA
 CC fragments were excised from the agarose gel, purified and ligated
 CC into the lambda ZAP II vector and probed with 31-A and 31-C. The
 CC phage clones were converted to the Bluescript SK(-) recombinant
 CC plasmid by in vivo excision. The recombinant expression of these
 CC rare OMP's can be used for diagnostic tests to detect syphilis and
 CC for development of host immunity during syphilis.
 XX Sequence 957 BP; 210 A; 206 C; 303 G; 238 T; 0 other;
 SQ

Query Match 75.2%; Score 15.8; DB 18; Length 957;
 Best Local Similarity 89.5%; Pred. No. 85;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 3 gtaactccctgctactct 21
 |||||||||||||
 Db 749 gtaactccctgctactct 767

RESULT 11
 AAA67251/C
 ID AAA67251 standard; DNA; 537 BP.
 XX AAA67251:
 AC 31-OCT-2000 (first entry)
 XX
 DT Eucalyptus grandis beta-amylase DNA sequence SEQ ID NO:252.
 DE
 XX
 XX Eucalyptus grandis; Pinus radiata; Monterey pine; modification;
 KM plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
 KM transgenic plant; ds.
 XX
 XX Eucalyptus grandis.
 OS
 XX WO200022092-A2.
 PN
 XX 20-APR-2000.
 PD
 XX 08-OCT-1999; 99WO-N200169.
 PF
 XX 13-OCT-1998; 98US-0170862.
 PR 11-AUG-1999; 99US-0148426.
 XX
 XX (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 PI
 XX Bloksberg LN;
 DR WPI; 2000-339328/29.
 DR
 XX New genes encoding proteins involved in a plant polysaccharide
 PT biosynthetic pathway, useful for modulating or altering the
 PT polysaccharide content, composition or structure of the plant
 XX
 PS Claim 1; Page 140; 301pp; English.
 XX The present invention describes isolated polynucleotides (PN) comprising
 CC a sequence selected from one of 835 nucleotide sequences given in
 CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
 CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,

CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
 CC 835 sequences or sequences that are degenerately equivalent or allelic
 CC to the 835 sequences. The polynucleotides are used to modify the
 CC activity of a polypeptide involved in a polysaccharide biosynthetic
 CC pathway in the plant. They are especially used to modulate or alter the
 CC polysaccharide content, composition or structure of the plant. AAB16268
 CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
 CC given in the present invention.

SO Sequence 537 BP; 155 A; 119 C; 136 G; 127 T; 0 other;

Query Match

Best Local Similarity 72.4%; Score 15.2; DB 21; Length 537;
 Pred. No. 1.6e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 agtgaactccctgactct 21

DB 472 AGTGAATCTCTCAACCTT 453

RESULT 12

ID AAV04426 standard; CDNA; 640 BP.

AC AAV04426;

DT 27-APR-1998 (first entry)

DE Human calpain small subunit cDNA (fragment).

KW Calpain; human; leukocyte; calcium dependent cysteine protease;
 KW screening; activator; inhibitor; treatment; prevention; cancer;
 KW cerebral apoplexy; cerebral infarction; subarachnoid haemorrhage;
 KW Alzheimer's disease; myodystrophy; cataracts; collagen disease;
 KW ischemic heart disease; atherosclerosis; arthritis;
 KW small subunit; ds.

OS Homo sapiens.

PN EP799892-A2.

PD 08-OCT-1997.

PE 03-APR-1997; 97EP-0105508.

PR 05-APR-1996; 96JP-0083649.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Kawamoto T, Nishi K, Shintani Y;

DR WPI: 1997-482674/45.

PT Human calpain protein and related DNA - useful for drug screening
 and treating cancer, stroke, etc.

PS Example 2; Page 36; 43pp; English.

CC The present sequence is a calpain small subunit cDNA fragment.
 CC Calpain is a human leukocyte derived calcium dependent cysteine
 CC protease. Calpain can be used to screen for compounds that activate
 CC or inhibit its proteolytic activity. Calpain DNA can be used to
 CC treat or prevent cancer, cerebral apoplexy, cerebral infarction,
 CC subarachnoid haemorrhage, Alzheimer's disease, myodystrophy,
 CC cataracts, ischemic heart disease, atherosclerosis, arthritis
 CC or collagen disease.

SO Sequence 640 BP; 139 A; 163 C; 223 G; 115 T; 0 other;

Query Match 72.4%; Score 15.2; DB 18; Length 640;
 Best Local Similarity 85.0%; Pred. No. 1.6e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 agtgaactccctgactct 21

DB 574 agtgaactccctgactct 593

RESULT 13

ID AAV04420 standard; CDNA; 804 BP.

AC AAV04420;

DT 27-APR-1998 (first entry)

DE Human calpain small subunit cDNA.

KW Calpain; human; leukocyte; calcium dependent cysteine protease;
 KW screening; activator; inhibitor; treatment; prevention; cancer;
 KW cerebral apoplexy; cerebral infarction; subarachnoid haemorrhage;
 KW Alzheimer's disease; myodystrophy; cataracts; collagen disease;
 KW ischemic heart disease; atherosclerosis; arthritis;
 KW small subunit; ds.

OS Homo sapiens.

PN EP799892-A2.

PD 08-OCT-1997.

PE 03-APR-1997; 97EP-0105508.

PR 05-APR-1996; 96JP-0083649.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Kawamoto T, Nishi K, Shintani Y;

DR WPI: 1997-482674/45.

DT P-PSDB; AAW41640.

PT Human calpain protein and related DNA - useful for drug screening
 and treating cancer, stroke, etc.

PS Disclosure; Page 35; 43pp; English.

CC The present sequence encodes calpain small subunit, a human
 CC leukocyte derived calcium dependent cysteine protease. Calpain can
 CC be used to screen for compounds that activate or inhibit its
 CC proteolytic activity. Calpain DNA can be used to treat or prevent
 CC cancer, cerebral apoplexy, cerebral infarction, subarachnoid
 CC haemorrhage, Alzheimer's disease, myodystrophy, cataracts,
 CC ischemic heart disease, atherosclerosis, arthritis or collagen
 CC disease.

SO Sequence 804 BP; 181 A; 203 C; 264 G; 156 T; 0 other;

Query Match 72.4%; Score 15.2; DB 18; Length 804;
 Best Local Similarity 85.0%; Pred. No. 1.7e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 agtgaactccctgactct 21

DB 574 agtgaactccctgactct 593

RESULT 14

AC7765
ID AAC7765 standard; cDNA, 1242 BP.

XX AAC7765;

XX 08-FEB-2001 (first entry)

XX Human cancer associated gene sequence SEQ ID NO:159.

XX Human; cancer associated gene; cancer antigen; detection; cancer;
XX diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
XX antidiabetic; antihistaminic; antirheumatic; antitumor; antiviral;
XX antineoplastic; antiproliferative; antitumor; antitumor; antitumor;
XX dermatological; neuroprotective; thrombolytic; coagulant; neutropenic;
XX vasotropic; antiproliferative; antitumor; antitumor; antitumor;
XX immune disorder; hematopoietic cell disorder; autoimmune disorder;
XX allergic reaction; graft versus host disease; organ rejection;
XX hematological; thrombolytic; cardiovascular disorder; infection;
XX neurological disease; drug screening; ss.

XX Homo sapiens.

XX WO20005350-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05882.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX NPI: 2000-587533/55.

XX P-PSDB: ABA43556.

XX Novel isolated nucleic acids comprising sequences encoding peptides
XX useful for treating or diagnosing e.g. cancer -

XX Claim 1: Page 737-738; 2352pp; English.

XX AAC77657 to AAC78448 encode the human cancer associated proteins given
XX in ABA43398 to ABA44239. The proteins can have activities based on the
XX tissues and cells the genes are expressed in. Example of activities
XX include: cytostatic; proliferative; vulnery; immunomodulator;
XX antidiabetic; antihistaminic; antirheumatic; antitumor; antitumor;
XX antineoplastic; antiproliferative; antitumor; antitumor; antitumor;
XX dermatological; neuroprotective; thrombolytic; coagulant;
XX vasotropic; antiproliferative; antitumor; antitumor; antitumor;
XX immune disorder; hematopoietic cell disorder; autoimmune disorder;
XX allergic reaction; graft versus host disease and organ
XX rejection, modulate haemostatic or thrombolytic activity, modulate
XX bacterial or viral infections. The peptides, nucleotides, antibodies,
XX agonists and antagonists may be also be used in drug screens. AAC78449 to
XX AAC78457 and ABA44240 represent sequences used in the exemplification of
XX the present invention.

XX Sequence 1242 BP; 333 A; 357 C; 292 G; 258 T; 2 other;

XX Query Match

XX 72.4%; Score 15.2; DB 21; Length 1242;

XX Best local similarity 85.0%; Pred. No. 1.8e-02;

XX Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

XX 2 aatgaactccctcactcctt 21

XX ||||||||| | | |||||

Db 414 aatgaactccctcactcctt 433

XX RESULT 15

XX AA298118/C

XX ID AA298118 standard; cDNA; 1741 BP.

XX AA298118;

XX 11-MAY-2000 (first entry)

XX Human signal peptide containing protein HSP-10 cDNA SEQ ID NO:144.

XX Human; signal peptide-containing protein; HSP; diagnosis; cancer;
XX inflammation; cardiovascular disease; anticancer; anti-inflammatory;
XX antitumor; neuroprotective; cardiovascular; hepatotropic;
XX antitumor; gene therapy; cell proliferation; neurological disorder;
XX reproductive disorder; developmental disorder; arteriosclerosis;
XX cirrhosis; psoriasis; acquired immune deficiency syndrome; anemia;
XX asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
XX Parkinson's disease; Huntington's disease; ovulatory defect;
XX muscular dystrophy; ss.

XX Homo sapiens.

XX WO20000610-A2.

XX 06-JAN-2000.

XX 25-JUN-1999; 99WO-US14484.

XX 26-JUN-1998; 98US-0090762.

XX 31-JUL-1998; 98US-0094983.

XX 01-OCT-1998; 98US-0102686.

XX 11-DEC-1998; 98US-0112129.

XX (INCYTE) INCYTE PHARM INC.

XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KI, Baughn MR;
XX Akredion IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
XX Bandman O;

XX NPI: 2000-150673/14.

XX P-PSDB: AAY7233.

XX New human signal peptide-containing proteins useful in treatment,
XX prevention and diagnosis of e.g. cancer, inflammation and
XX cardiovascular disease -

XX Claim 9: Page 257-258; 327pp; English.

XX AA298109 to AA298242 encode AAY7224 to AAY7357 which represent the
XX human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have
XX anticancer, anti-inflammatory, antimicrobial, neurotropic, hepatotropic,
XX neuroprotective, cardiovascular and antitumor activities, and can
XX be used in gene therapy. HSPs can be used to treat or prevent disorders
XX associated with decreased activity or function of HSP. Antagonists of
XX HSP are used to treat or prevent disorders associated with increased
XX activity or function of HSP. Such disorders include cell proliferation
XX (including cancer), inflammation, cardiovascular, neurological,
XX reproductive or developmental disorders, (e.g. arteriosclerosis,
XX cirrhosis, psoriasis, acquired immune deficiency syndrome, anemia,
XX asthma, Crohn's disease, microbial or other infections, congestive or
XX ischemic heart disease, Alzheimer's, Parkinson's or Huntington's
XX diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP
XX nucleic acids can be used for the recombinant production of HSP, for
XX detecting HSP in standard hybridisation and amplification assays (for
XX diagnosis and monitoring), in gene therapy, as antisense,
XX triplex-forming or ribozyme therapeutics, for detecting related sequences
XX or genetic variations, and for chromosomal mapping. HSP are also used to
XX raise specific antibodies (Ab) and to screen for agonists and
XX antagonists (potential therapeutic agents). Ab are used to diagnose, or
XX monitor, HSP-related diseases (in usual immunoassays), as therapeutic

CC antagonists, in competitive drug screens, and for purification of HSP
CC from natural sources.

XX
SQ Sequence 1741 BP; 533 A; 358 C; 364 G; 486 T; 0 other;

Query Match 72.4%; Score 15.2; DB 21; Length 1741;
Best Local Similarity 85.0%; Pred. No. 1.8e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 aagtgactcccgctact 20
|||||

Db 1411 AAGTGAAGTGTCTCTACT 1392

Search completed: July 25, 2001, 05:23:03
Job Time: 4676 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 02:45:17 ; Search time 117.39 seconds
(without alignments)
33.141 Million cell updates/sec

Title: US-09-142-095-1

Sequence: 1 aagtagctccctcctcctcct 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summariesDatabase : Issued Patents, NA.*
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	15.8	75.2	957 1	US-08-599-480-1
2	15.8	75.2	957 1	US-08-842-198-1
3	15.8	75.2	957 5	PCT-US85-00180-1
4	15.2	72.4	640 2	US-08-835-099A-16
5	15.2	72.4	640 3	US-09-157-349-16
6	15.2	72.4	804 2	US-08-835-099A-10
7	15.2	72.4	804 3	US-09-157-349-10
8	15.2	72.4	804 3	US-08-724-394A-20
9	15.2	72.4	246240 2	US-08-724-394A-21
10	15.2	72.4	246240 2	US-08-724-394A-22
11	14.8	70.5	1723 1	US-08-417-492-1
12	14.8	70.5	2161 2	US-09-001-826-4
13	14.8	70.5	2161 2	US-09-001-826-25
14	14.8	70.5	2652 2	US-08-366-547-1
15	14.8	70.5	2977 1	US-08-276-151-1
16	14.8	70.5	2977 1	US-08-306-691B-17
17	14.8	70.5	2977 1	US-08-756-806A-64
18	14.8	70.5	2977 3	US-08-328-239A-4
19	14.8	70.5	2977 3	US-09-143-214-64
20	14.8	70.5	2977 3	US-09-209-668-12
21	14.8	70.5	2977 5	PCT-US95-13661-4
22	14.6	69.5	394 1	US-08-488-144-18
23	14.6	69.5	1168 4	US-08-930-285-15
24	14.6	69.5	4345 2	US-08-244-537-1
25	14.6	69.5	7470 3	US-08-417-089-5
26	14.6	69.5	7470 4	US-08-695-651-5
27	14.6	69.5	7470 4	US-08-930-285-5

28	14.4	68.6	1273 4	US-08-725-758A-3	Sequence 3, Appl
29	14.4	68.6	1373 4	US-08-725-758A-1	Sequence 1, Appl
30	14.4	68.6	4310 4	US-09-008-172-1	Sequence 1, Appl
31	14.2	67.6	61 4	US-08-687-421-355	Sequence 355, App
32	14.2	67.6	1118 1	US-08-418-032-1	Sequence 1, Appl
33	14.2	67.6	1554 1	US-08-469-486-1	Sequence 1, Appl
34	14.2	67.6	1554 2	US-08-469-486-1	Sequence 1, Appl
35	14.2	67.6	3504 1	US-08-620-717A-8	Sequence 8, Appl
36	14.2	67.6	5541 1	US-08-920-812-20	Sequence 20, Appl
37	14.2	67.6	5541 1	US-08-920-827-20	Sequence 20, Appl
38	14.2	67.6	5541 1	US-08-921-177-20	Sequence 20, Appl
39	14.2	67.6	5541 1	US-08-920-812-20	Sequence 20, Appl
40	14.2	67.6	5541 1	US-08-362-577C-20	Sequence 20, Appl
41	14.2	67.6	5541 2	US-08-920-828-20	Sequence 20, Appl
42	14.2	67.6	20303 1	US-08-370-975B-6	Sequence 20, Appl
43	14.2	67.6	26764 1	US-08-370-975B-1	Sequence 20, Appl
44	14.2	67.6	1441 1	US-08-136-277-18	Sequence 18, Appl
45	14.2	67.6	1441 2	US-08-479-403-18	Sequence 18, Appl
			1441 3	US-08-835-734-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-599-480-1
Sequence 1, Application US/08599480
Patent No. 5753459
GENERAL INFORMATION:
APPLICANT: Blanco, David R.
APPLICANT: Miller, James N.
APPLICANT: Lovett, Michael A.
APPLICANT: Champion, Cheryl I.
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF A
TITLE OF INVENTION: T. Pallidum RARE OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/599,480
FILING DATE: 23-JAN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Learn, June M.
REGISTRATION NUMBER: 31,238
TELEPHONE/DOCKET NUMBER: 07419/018001 (CIP of 016001)
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 957 base pairs
TYPE: nucleic acid
STANDARDS: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: TROMP1
FEATURES:
NAME/KEY: CDS
LOCATION: 1..954
US-08-599-480-1

Query Match 75.2%; Score 15.8; DB 1; Length 957;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 gtgaactccctgctactctt 21
|||
Db 749 GTAACCTCCCTGCTACTT 767

RESULT 2
US-08-842-199-1
Sequence 1, Application US/08842199
Patent No. 5821085
GENERAL INFORMATION:
APPLICANT: Bianco, David R.
APPLICANT: Miller, James N.
APPLICANT: Lovett, Michael A.
APPLICANT: Champion, Cheryl I.
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF A
TITLE OF INVENTION: T. Pallidum RARE OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Judas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/842,199
FILING DATE: 23-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/292,904
FILING DATE: 17-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Boslich, June M.
REGISTRATION NUMBER: 31,238
REFERENCE/DOCKET NUMBER: PD-3744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 957 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: TROMP1
FEATURE:
NAME/KEY: CDS
LOCATION: 1..954
US-08-842-199-1

Query Match 75.2%; Score 15.8; DB 1; Length 957;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 gtgaactccctgctactctt 21
|||
Db 749 GTAACCTCCCTGCTACTT 767

RESULT 3

PCT-US95-00190-1
Sequence 1, Application PC/TUS9500190
GENERAL INFORMATION:
APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF A
TITLE OF INVENTION: T. Pallidum RARE OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Judas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00190
FILING DATE: 06-JAN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Boslich, June M.
REGISTRATION NUMBER: 31,238
REFERENCE/DOCKET NUMBER: PD-3744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 957 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: TROMP1
FEATURE:
NAME/KEY: CDS
LOCATION: 1..954
PCT-US95-00190-1

Query Match 75.2%; Score 15.8; DB 5; Length 957;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 gtgaactccctgctactctt 21
|||
Db 749 GTAACCTCCCTGCTACTT 767

RESULT 4
US-08-835-099A-16
Sequence 15, Application US/08835099A
Patent No. 5874277
GENERAL INFORMATION:
APPLICANT: SHINTANI, Yasushi
APPLICANT: NISHI, Kazuo
APPLICANT: KAWAMOTO, Tomohiro
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
TITLE OF INVENTION: AND USE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/835,099A
FILING DATE: 04-APR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 083649/1996
FILING DATE: 05-APR-1996
APPLICATION NUMBER: 97105508.2
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 640 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-835-099A-16

Query Match 72.4% Score 15.2; DB 2; Length 640;
Best Local Similarity 85.0%; Pred. No. 36;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 agtgaactccctgctaccct 21
|||||
Db 574 agtgaactccctgctaccct 593

RESULT 5
US-09-157-349-16
Sequence 16, Application US/09157349
Patent No. 6068990
GENERAL INFORMATION:
APPLICANT: SHINTANI, Yasushi
APPLICANT: NISHI, Kazuo
APPLICANT: KAWAMOTO, Tomohiro
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
TITLE OF INVENTION: AND USE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/157,349
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/835,099
FILING DATE:
APPLICATION NUMBER: 97105508.2
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:

NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 640 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-157-349-16

Query Match 72.4% Score 15.2; DB 3; Length 640;
Best Local Similarity 85.0%; Pred. No. 36;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 agtgaactccctgctaccct 21
|||||
Db 574 agtgaactccctgctaccct 593

RESULT 6
US-08-835-099A-10
Sequence 10, Application US/08835099A
Patent No. 5874277
GENERAL INFORMATION:
APPLICANT: SHINTANI, Yasushi
APPLICANT: NISHI, Kazuo
APPLICANT: KAWAMOTO, Tomohiro
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
TITLE OF INVENTION: AND USE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/835,099A
FILING DATE: 04-APR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 083649/1996
FILING DATE: 05-APR-1996
APPLICATION NUMBER: 97105508.2
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 804 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA

US-08-835-099A-10

Query Match 72.4%; Score 15.2; DB 2; Length 804;
Best Local Similarity 85.0%; Pred. No. 37;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 agtgaactccgtctacct 21
DB 574 AGTGACTCCAGTGCTT 593

RESULT 7

US-09-157-349-10
Sequence 10, Application US/09157349
Patent No. 6068990

GENERAL INFORMATION:

APPLICANT: SHINTANI, Yasushi

APPLICANT: NISHI, Kazuo

APPLICANT: KAMAMOTO, Tomohito

TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/157,349

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/835,099

FILING DATE:

APPLICATION NUMBER: 97105508.2

FILING DATE: 03-APR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Resnick, David S

REGISTRATION NUMBER: 34,235

REFERENCE/DOCKET NUMBER: 47342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

TELEX:

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 804 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-09-157-349-10

Query Match

Best Local Similarity 72.4%; Score 15.2; DB 3; Length 804;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 agtgaactccgtctacct 21
DB 574 AGTGACTCCAGTGCTT 593

RESULT 8

US-08-724-394A-20/C

Sequence 20, Application US/08724394A
Patent No. 587237

GENERAL INFORMATION:

APPLICANT: Feder, John N.

APPLICANT: Kronmal, Gregory S.

APPLICANT: Lauer, Peter M.

APPLICANT: Ruddy, David A.

APPLICANT: Thomas, Winston

APPLICANT: Tsuchihashi, Zenta

APPLICANT: Wolff, Roger K.

TITLE OF INVENTION: Megabase Transcript Map: No. 587237el

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND and CREW LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724,394A

FILING DATE: 01-OCT-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Fitts, Renee A.

REGISTRATION NUMBER: 35,136

REFERENCE/DOCKET NUMBER: 017957-000100

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-576-0200

TELEFAX: 415-576-0300

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 246240 base pairs

TYPE: nucleic acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..246240

OTHER INFORMATION: /note="HLA-H.CONTIG"

US-08-724-394A-20

Query Match

Best Local Similarity 72.4%; Score 15.2; DB 2; Length 246240;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 aagtgaactccgtctacct 20
DB 139464 ATGTGACTCCTCACTT 139445

RESULT 9

US-08-724-394A-21/C
Sequence 21, Application US/08724394A
Patent No. 587237

GENERAL INFORMATION:

APPLICANT: Feder, John N.

APPLICANT: Kronmal, Gregory S.

APPLICANT: Lauer, Peter M.

APPLICANT: Ruddy, David A.

APPLICANT: Thomas, Winston

APPLICANT: Tsuchihashi, Zenta

APPLICANT: Wolff, Roger K.

TITLE OF INVENTION: Megabase Transcript Map: No. 587237el

Sequences and Antibodies Thereeto

NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Pitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note="HLA-H.CONTIG"
US-08-724-394A-21

Query Match 72.4%; Score 15.2; DB 2; Length 246240;
Best Local Similarity 85.0%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 aagtgactccctgactact 20
Db 139464 ATGTGACTGCTGCAACT 139445

RESULT 10
US-08-724-394A-22/C
Sequence 22, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolf, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Pitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note="HLA-H.CONTIG"
US-08-724-394A-22

Query Match 72.4%; Score 15.2; DB 2; Length 246240;
Best Local Similarity 85.0%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 aagtgactccctgactact 20
Db 139464 ATGTGACTGCTGCAACT 139445

RESULT 11
US-08-417-492-1
Sequence 1, Application US/08417492
Patent No. 5750872
GENERAL INFORMATION:
APPLICANT: Bennett, Alan B.
APPLICANT: Brunnett, David A.
APPLICANT: Grunltz, Alexander A.
TITLE OF INVENTION: Nucleic Acids Encoding Ascorbate Free
TITLE OF INVENTION: Radical Reductase and Their Uses
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew
STREET: One Market Plaza, Stewart Street Tower
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105-1492
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/417,492
FILING DATE: 05-APR-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 2307E-586US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-543-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1723 base pairs
TYPE: nucleic acid

```
STRANDEDNESS: single
TOPOLOGY: linear
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: exon
LOCATION: 49..643
OTHER INFORMATION: /number- 1
FEATURE:
NAME/KEY: intron
LOCATION: 644..780
OTHER INFORMATION: /number- 1
FEATURE:
NAME/KEY: exon
LOCATION: 781..1484
OTHER INFORMATION: /number- 2
FEATURE:
NAME/KEY: CDS
LOCATION: join(49..643, 781..1487)
US-08-417-492-1

Query Match
Best Local Similarity 70.5%; Score 14.8; DB 1; Length 1723;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tgaactccctgatacctt 21
DB 1424 TCACCCCTCTCTACTT 1441

RESULT 12
US-09-001-826-4
; Sequence 4, Application US/09001826A
; Patent No. 5965727
; GENERAL INFORMATION:
; APPLICANT: SONG, HEE-SOOK
; APPLICANT: BROTHERTON, JEFFREY E.
; APPLICANT: MIDHOLM, JACK M.
; TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE
; FILE REFERENCE: 07001.C1
; CURRENT APPLICATION NUMBER: US/09/001,826A
; CURRENT FILING DATE: 1997-12-31
; EARLIER APPLICATION NUMBER: 08/937,739; 60/025,140
; EARLIER FILING DATE: 1997-07-25; 1996-07-26
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text
; SEQ ID NO 4
; LENGTH: 2161
; TYPE: DNA (cDNA)
; ORGANISM: Nicotiana tabacum
US-09-001-826-4

Query Match
Best Local Similarity 70.5%; Score 14.8; DB 2; Length 2161;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tgaactccctgatacctt 21
DB 593 tgaactccctgatacctt 610

RESULT 13
US-09-001-826-25
; Sequence 25, Application US/09001826A
; Patent No. 5965727
; GENERAL INFORMATION:
; APPLICANT: SONG, HEE-SOOK
; APPLICANT: BROTHERTON, JEFFREY E.
; APPLICANT: MIDHOLM, JACK M.
; TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE
; FILE REFERENCE: TRANSFORMATION
```

```
FILE REFERENCE: 07001.C1
; CURRENT APPLICATION NUMBER: US/09/001,826A
; CURRENT FILING DATE: 1997-12-31
; EARLIER APPLICATION NUMBER: 08/937,739; 60/025,140
; EARLIER FILING DATE: 1997-07-25; 1996-07-26
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text
; SEQ ID NO 25
; LENGTH: 2161
; TYPE: DNA (cDNA)
; ORGANISM: Nicotiana tabacum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..1940
US-09-001-826-25

Query Match
Best Local Similarity 70.5%; Score 14.8; DB 2; Length 2161;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tgaactccctgatacctt 21
DB 593 tgaactccctgatacctt 610

RESULT 14
US-08-366-547-1
; Sequence 1, Application US/08366547
; Patent No. 5843737
; GENERAL INFORMATION:
; APPLICANT: Chen, Ian Bo
; APPLICANT: Bao, Shideng
; TITLE OF INVENTION: A NEW CANCER ASSOCIATED GENE, PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS
; ADDRESS: 6 CUSHMAN
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/366,547
; FILING DATE: 30-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Eisenstein, Ronald I.
; REGISTRATION NUMBER: 30628
; REFERENCE/DOCKET NUMBER: 45072
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2652 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 120..2130
US-08-366-547-1
```

Query Match 70.5%; Score 14.8; DB 2; Length 2652;
 Best Local Similarity 88.9%; Pred. No. 72;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 agtgactccctgctac 19
 |||||
 Db 2037 AGTGAAGCTGCTGCTAC 2054

RESULT 15

US-08-276-151-1/c
 ; Sequence 1, Application US/08276151
 ; Patent No. 5597719

GENERAL INFORMATION:

APPLICANT: Freed, Ellen
 APPLICANT: Ruggieri, Rosamaria
 TITLE OF INVENTION: Interaction of raf-1 and 14-3-3 proteins
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooley Godward et al.
 STREET: Five Palo Alto Square
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA

ZIP: 94036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/276,151
 FILING DATE: 14-JUL-1994

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Torchia, Ph.D., Timothy E

REGISTRATION NUMBER: 36,700

REFERENCE/DOCKET NUMBER: ONTX-005/0005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 843-5481

TELEFAX: (415) 857-0663

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2977 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

HYPOTHEICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: 130..2076

US-08-276-151-1

Query Match 70.5%; Score 14.8; DB 1; Length 2977;
 Best Local Similarity 88.9%; Pred. No. 74;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 4 tgaactccctgctactt 21
 |||||
 Db 2436 TGGAAGCTGCTGCTACTT 2419

Search completed: July 25, 2001, 05:18:51
 Job time: 3214 sec

QY 3 tgcctcgcagaggtt 19
Db 506 tgcctcgcagaggtt 490

RESULT 5

LOCUS A0068059

DEFINITION HS_2200_B2_F03_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2200 Col-6 Row-L, DNA sequence.

ACCESSION

A0068059

VERSION A0068059.1 GI:3379327

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 356)

AUTHORS

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kellier,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE

99380589

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2200 row: L column: 6
Class: BAC ends
High quality sequence stop: 356.

FEATURES

Location/Qualifiers
1..356
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_plate="Plate-2200 Col-6 Row-L"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBAC3; BAC Clones in E-Coll DH10B"

BASE COUNT

111 a 86 c 67 g 92 t

ORIGIN

Query Match 86.3%; Score 16.4; DB 223; Length 356;
Best Local Similarity 94.4%; Pred. NO. 6.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

2 tgcctcgcagaggtt 19

Db

106 tgcctcgcagaggtt 123

RESULT 6

LOCUS A0799987

DEFINITION

HS_5300_B2_G10_T7A RPCT-11 Human Male BAC Library Homo sapiens genomic clone Plate-876 Col-20 Row-N, DNA sequence.

ACCESSION

A0799987

VERSION

A0799987.1 GI:5717319

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 530)

AUTHORS

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kellier,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE

99380589

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 876 row: N column: 20
Class: BAC ends
High quality sequence stop: 530.

FEATURES

Location/Qualifiers
1..530
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_plate="Plate-876 Col-20 Row-N"
/clone_lib="RPCT-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3; 6; Site:1; EcoRI; Site:2; EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

BASE COUNT

109 a 120 c 167 g 130 t

ORIGIN

Query Match 86.3%; Score 16.4; DB 233; Length 530;
Best Local Similarity 94.4%; Pred. NO. 6.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1 tgcctcgcagaggtt 18

Db

325 tgcctcgcagaggtt 342

RESULT 7

LOCUS A1979827

DEFINITION

A1979827 574 bp mRNA EST 10-JUL-2000
pat.PK0004.f2 chicken activated T cell cDNA Gallus gallus CDNA clone pat.PK0004.f2 5' similar to FRUCTOSE-BISPHOSPHATE ALDOLASE B (LIVER-TYPE ALDOLASE), mRNA sequence.

ACCESSION

A1979827

VERSION

A1979827.1 GI:5882855

KEYWORDS

EST.

SOURCE

chicken.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE

1 (bases 1 to 574)

AUTHORS

Thirunagaru,V.G., Sofer,L., Cul,J. and Burnside,J.
An expressed sequence tag database of T-cell-enriched activated chicken splenocytes: sequence analysis of 5251 clones
Genomics 66 (2), 144-151 (2000)

CONTACT

Contact: Joan Burnside
Molecular Endocrinology
University of Delaware
40 Townsend Hall, Newark, DE 19717, USA
Tel: 302 831-1345
Fax: 302 831-3651

Email: joan@udel.edu
Seq primer: T7.

FEATURES
source
Location/Qualifiers

1. 574
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone_pat="PK0004.12"
/clone_lib="chicken activated T cell cDNA"
/sex="male"
/cell_type="Con A-activated splenic T cell"
/lab_host="E.coli TOP10 F"
/note="Vector: pCDNA3"

BASE COUNT
ORIGIN
165 a 130 c 154 g 125 t

Query Match
Best Local Similarity 94.4%; Score 16.4; DB 104; Length 574;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ttgtctctgcagaggt 18
|||||
Db 340 TTGTCCTGCGCAGAGCT 323

RESULT 8
LOCUS A2664475 616 bp DNA GSS 14-DEC-2000
DEFINITION IM0544H08R Mouse 10kb plasmid UUGCM library Mus musculus genomic
clone UUGCM0544H08 R, DNA sequence.
ACCESSION A2664475
VERSION A2664475.1 GI:11801621
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 616)
Dunn, D., Moyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0544 row: H column: 08
Seq primer: CACACAGAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 616.

FEATURES

source

Location/Qualifiers

1. 616

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGCM0544H08"

/clone_lib="Mouse 10kb plasmid UUGCM library"

/sex="Male"

/lab_host="E. coli strain XL10-gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA

BASE COUNT
ORIGIN
156 a 162 c 131 g 167 t

Query Match
Best Local Similarity 94.4%; Score 16.4; DB 247; Length 616;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ttgtctctgcagaggt 19
|||||
Db 319 TTGTCCTGCGCAGAGCTT 336

RESULT 9
LOCUS BE901069 680 bp mRNA EST 29-SEP-2000
DEFINITION 601674415P1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3957315 5',
mRNA sequence.
ACCESSION BE901069
VERSION BE901069.1 GI:10389877
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 680)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: ccaps-f@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: LCN834 row: n column: 04
High quality sequence start: 2
High quality sequence stop: 608.

FEATURES

source

Location/Qualifiers

1. 680

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3957315"

/clone_lib="NIH_MGC_21"

/tissue_type="choriocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: placenta; Vector: pOTB; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT
ORIGIN
117 a 190 c 214 g 159 t

Query Match 86.3%; Score 16.4; DB 141; Length 680;
 Best Local Similarity 94.4%; Pred. No. 6.9e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ttgtctcctccagaggt 18
 DB 371 TTGTCTCCTCCAGAGGT 388

RESULT 10
 BG109825/c 978 bp mRNA EST 30-JAN-2001
 LOCUS 60280983F1 NIH_MGC-86 Homo sapiens cDNA clone IMAGE:4368405 5',
 DEFINITION mRNA sequence.
 ACCESSION BG109825.1 GI:12603331
 VERSION EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 978)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LAM10022 row: f column: 22
 High quality sequence stop: 751.
 Location/Qualifiers
 1. 978
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4368405"
 /clone_lib="NIH_MGC-86"
 /tissue_type="osteosarcoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: bone; Vector: PCMV-SPORT6, Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.533 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH-MGC Library."

BASE COUNT 250 a 250 c 276 g 202 t
 ORIGIN

Query Match 86.3%; Score 16.4; DB 173; Length 978;
 Best Local Similarity 94.4%; Pred. No. 7.3e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ttgtctcctccagaggt 18
 DB 844 TTGTCTCCTCCAGAGGT 827

RESULT 11
 AI203923 206 bp mRNA EST 28-OCT-1998
 LOCUS 3', mRNA sequence.
 DEFINITION qd7h07.x1 Soares, testis_NHT Homo sapiens cDNA clone IMAGE:1735069
 ACCESSION AI203923
 VERSION AI203923.1 GI:3756529
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 206)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
 , Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbdp/image/image.html
 Insert length: 1344 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 197.
 Location/Qualifiers
 1. 206
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1735069"
 /clone_lib="Soares-testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5'
 TGTTCACATCTGATGAGGACGGCCGCCCAATTTTCTTTTCTTTT 3']
 (Pharmacia), digested with Not I and Eco RI adaptors
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization to Cols5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 50 a 48 c 53 g 55 t
 ORIGIN

Query Match 84.2%; Score 16; DB 17; Length 206;
 Best Local Similarity 100.0%; Pred. No. 9.1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 gctcctccagaggt 19
 DB 52 GCTCCTCCAGAGGT 67

RESULT 12
 AA504455 338 bp mRNA EST 18-AUG-1997
 LOCUS aa60a05.s1 NCI-CGAP-GCB1 Homo sapiens cDNA clone IMAGE:825296 3',
 DEFINITION mRNA sequence.
 ACCESSION AA504455
 VERSION AA504455.1 GI:2240615
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 338)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 320.
 Location/Qualifiers

FEATURES

source
 1. 338
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:825296"
 /clone_1lb="NCI-CGAP-CGB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: p773D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20⁺, IgD⁺),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
 primed with a Not I - 0190(dT) primer
 [5'-TGTACCAATCGAAGTGGAGGCGCGCTCTTTTCTTTTCTTTT-3',
 1. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p773 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 79 a 71 c 68 g 120 t
 ORIGIN

Query Match 84.2%; Score 16; DB 8; Length 338;
 Best Local Similarity 100.0%; Pred. No. 9.9e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gctctgcagaggtt 19
 |||||
 Db 86 gctctgcagaggtt 101

RESULT 13
 AA577198 341 bp MRNA EST 12-SEP-1997
 LOCUS m87b11.s1 NCI-CGAP-co9 Homo sapiens cDNA clone IMAGE:1075197 3'
 DEFINITION MRNA sequence.
 AA577198
 VERSION AA577198.1 GI:2354672
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 341)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
 Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Insert length: 919 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 181.
 Location/Qualifiers

FEATURES

source
 1. 341
 /organism="Homo sapiens"

/db_xref="taxon:9606"
 /clone="IMAGE:1075197"
 /clone_1lb="NCI-CGAP-co9"
 /tissue_type="colon tumor RER"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: p773D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 RER+ colon tumor, and was then primed with a Not I -
 0190(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified p773
 vector. Library is not normalized. Library was
 constructed by Bento Soares and M. Fatima Bonaldo (Soares4
)."

BASE COUNT 79 a 85 c 82 g 95 t
 ORIGIN

Query Match 84.2%; Score 16; DB 9; Length 341;
 Best Local Similarity 100.0%; Pred. No. 9.9e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gctctgcagaggtt 19
 |||||
 Db 189 gctctgcagaggtt 204

RESULT 14
 A0751483 738 bp DNA GSS 19-JUL-1999
 LOCUS HS-5576_p2-B02_Sp6 RPCI-11 Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate-1152 Col-4 Row-D, DNA sequence.
 A0751483
 VERSION A0751483.1 GI:5538641
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 738)
 AUTHORS Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
 Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
 Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams K.D. and
 Hood L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas G.G., Wallace J.C., Hood L.
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 or from Research Genetics (<http://inforesgen.com>). BAC end Web Server:
<http://www.htsc.washington.edu>
 Plate: 1152 Row: D Column: 4
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 738.
 Location/Qualifiers

FEATURES

source
 1. 738
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="plate-1152 Col-4 Row-D"
 /clone_1lb="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor

and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"

BASE COUNT

178 a 167 c 190 g 203 t

OY 1 ttgtccctgcagaggtt 19
DB 49 ttgtccctgcagaggtt 31

Query Match

84.2%; Score 16; DB 233; Length 728;

Best Local Similarity 100.0%; Pred. No. 1.1e+03;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gtctccctgcagaggtt 19

DB 197 ctctccctgcagaggtt 212

RESULT 15

W25430/c

LOCUS

DEFINITION

W25430

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

W25430 115 bp mRNA EST 10-OCT-1996
2590c01.r1 Soares, senescent_fibroblasts_NBHSF Homo sapiens cDNA
clone IMAGE:310848 5', mRNA sequence.
W25430
W25430.1 GI:1303285
EST
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
Ritkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston
R., Williamson, A., Wohlmann, P., and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 627 Std Error: 0.00
Seq primer: mob.REGA-ET.
Location/Qualifiers
1. 115
/organism="Homo sapiens"
/db_xref="GDB:1254262"
/db_xref="taxon:9606"
/clone_id="Soares_senescent_fibroblasts_NBHSF"
/tissue_type="senescent_fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pTZ19 (Pharmacia) with a modified
polylinker V-type: phagemid; site 1: Not I; site 2: Eco RI
; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTACCACTGAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pTZ19 vector
(Pharmacia). Library went through one round of
normalization to a Cot - 5. Library constructed by Bento
Soares and M. Fatima Bonaldo."

FEATURES

SOURCE

BASE COUNT

40 a 17 c 27 g 31 t

Query Match

83.2%; Score 15.8; DB 190; Length 115;

Best Local Similarity

89.5%; Pred. No. 1.1e+03;

Matches 17; Conservative

0; Mismatches 2; Indels 0; Gaps 0;

Search completed: July 25, 2001, 04:54:56
Job time: 10429 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:23:04 ; Search time 247.87 Seconds
(without alignments)
50.664 Million cell updates/sec

Title: US-09-142-095-3

Perfect score: 20
Sequence: 1 gtcacgtgacacgtcaac 20

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

N_Geneseq_0601:*

1:	/SIDSL/gcgdata/geneseq/NA1980.DAT:*
2:	/SIDSL/gcgdata/geneseq/NA1981.DAT:*
3:	/SIDSL/gcgdata/geneseq/NA1982.DAT:*
4:	/SIDSL/gcgdata/geneseq/NA1983.DAT:*
5:	/SIDSL/gcgdata/geneseq/NA1984.DAT:*
6:	/SIDSL/gcgdata/geneseq/NA1985.DAT:*
7:	/SIDSL/gcgdata/geneseq/NA1986.DAT:*
8:	/SIDSL/gcgdata/geneseq/NA1987.DAT:*
9:	/SIDSL/gcgdata/geneseq/NA1988.DAT:*
10:	/SIDSL/gcgdata/geneseq/NA1989.DAT:*
11:	/SIDSL/gcgdata/geneseq/NA1990.DAT:*
12:	/SIDSL/gcgdata/geneseq/NA1991.DAT:*
13:	/SIDSL/gcgdata/geneseq/NA1992.DAT:*
14:	/SIDSL/gcgdata/geneseq/NA1993.DAT:*
15:	/SIDSL/gcgdata/geneseq/NA1994.DAT:*
16:	/SIDSL/gcgdata/geneseq/NA1995.DAT:*
17:	/SIDSL/gcgdata/geneseq/NA1996.DAT:*
18:	/SIDSL/gcgdata/geneseq/NA1997.DAT:*
19:	/SIDSL/gcgdata/geneseq/NA1998.DAT:*
20:	/SIDSL/gcgdata/geneseq/NA1999.DAT:*
21:	/SIDSL/gcgdata/geneseq/NA2000.DAT:*
22:	/SIDSL/gcgdata/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	18	AAT79543
2	20	100.0	620	18	AAT79540
3	16.4	82.0	22	19	AAV13057
4	15.8	79.0	27	21	AAZ57627
5	15.8	79.0	27	21	AAZ48039
6	15.8	79.0	300	20	AAZ98515
7	15.8	79.0	392	21	AAZ50369
8	15.8	79.0	507	21	AAZ50370
9	15.8	79.0	680	21	AAZ50368
10	15.8	79.0	685	21	AAZ50367
11	15.8	79.0	738	21	AAZ57158

12	15.8	79.0	900	20	AAZ20091
13	15.8	79.0	900	21	AAZ80993
14	15.2	76.0	1276	21	AAZ13921
15	15.2	76.0	1801	21	AAZ35997
16	15.2	76.0	1803	21	AAZ31372
17	15	75.0	2530	21	AAZ39424
18	14.8	74.0	393	22	AAZ66883
19	14.8	74.0	430	16	AAZ19914
20	14.8	74.0	519	19	AAZ69028
21	14.8	74.0	519	21	AAZ80982
22	14.8	74.0	887	7	AAZ60894
23	14.8	74.0	1100	5	AAZ0127
24	14.8	74.0	1100	5	AAZ0103
25	14.8	74.0	1101	4	AAZ30003
26	14.8	74.0	1101	5	AAZ40133
27	14.8	74.0	1101	5	AAZ40240
28	14.8	74.0	1144	4	AAZ30109
29	14.8	74.0	1156	6	AAZ03276
30	14.8	74.0	1161	6	AAZ0327
31	14.8	74.0	1161	7	AAZ0103
32	14.8	74.0	1210	4	AAZ30055
33	14.8	74.0	1226	13	AAZ24919
34	14.8	74.0	1350	6	AAZ50335
35	14.8	74.0	2059	7	AAZ60895
36	14.8	74.0	2171	22	AAZ54722
37	14.8	74.0	2274	22	AAZ54721
38	14.8	74.0	2740	16	AAZ085355
39	14.8	74.0	2749	20	AAZ08488
40	14.8	74.0	2827	7	AAZ60893
41	14.8	74.0	3000	7	AAZ60892
42	14.8	74.0	5274	21	AAZ42949
43	14.8	74.0	5319	21	AAZ50069
44	14.8	74.0	5961	16	AAZ0060
45	14.8	74.0	68940	20	AAZ57351

ALIGNMENTS

RESULT 1					
ID	AAZ79543	standard; DNA; 20 BP.			
AC	AAZ79543;				
DE	23-JAN-1998	(first entry)			
XX	UGT1*1 gene exon 1 upstream PCR primer C.				
KW	Uridine diphosphate glucuronosyltransferase gene; UGT1;				
KW	Gilbert's syndrome; GS; unconjugated hyperbilirubinemia;				
KW	bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;				
KW	Drug trial efficiency; screening; PCR primer; ss.				
XX	Synthetic.				
OS	Homo sapiens.				
XX					
PN	W09732042-A2.				
XX					
PD	04-SEP-1997.				
XX					
PF	03-MAR-1997; 97MO-GB00577.				
XX					
PR	16-MAR-1996; 96GB-0005598.				
XX					
PA	01-MAR-1996; 96GB-0004480.				
XX					
PI	(UTD-) UNIV DUNDEE.				
XX					
PI	Burchell B;				
XX					
DR	WPI: 1997-448702/41.				
XX					
PT	Improving drug trial efficiency comprises identifying participants				

Type IV collagen N
Human alpha3(IV)NC
Aspergillus oryzae
Arabidopsis thaliana
Arabidopsis thaliana
Arabidopsis thaliana
Novel human polyn
Human gene signatu
DNA molecule encod
Human breast tumor
Sequence of plasmid
Sequence encoding
Sequence of human
Sequence including
Sequence of human
Human immunointerf
Sequence encoding
Human recombinant
Human interferon-g
Human interferon-g
Sequence of the p1
Human interferon-g
Sequence of the p1
Nucleotide sequenc
Nucleotide sequenc
Prosoposin cDNA.
Human prosoposin c
Sequence of plasmid
Arabidopsis thaliana
Arabidopsis thaliana
Human interferon-g
Human chromosome 6

PT with Gilbert's syndrome - useful as their altered drug metabolism
 PT may hinder result interpretation
 PS Claim 14; Page 12; 31pp; English.

CC This PCR primer (with primers AAT79542-44) flanks the TATA box sequence
 CC upstream of the uridine diphosphate glucuronosyltransferase (UGT) gene
 CC 11 exon 1 (see AAT79540), and was used to amplify fragments of 98-100
 CC bp. This gene is known to be associated with Gilbert's syndrome (GS). GS
 CC is a mild, common form of unconjugated hyperbilirubinaemia associated
 CC with reduced bilirubin glucuronidation capacity. Analysis of the genetic
 CC basis of GS has allowed 2 forms to be identified. One is a mild form
 CC upstream of the UGT1*1 exon 1, and the other is a more severe form
 CC associated with heterozygosity for a mutation which, when homozygous,
 CC causes Crigler-Najjar type 2 disease. The first form is autosomal
 CC recessive and the second is inherited dominantly. Patients suffering from
 CC GS, which is benign, may have altered metabolism of some drugs, making it
 CC difficult to determine if an effect is due to the drug or the syndrome.
 CC Drug trial efficiency would be improved if potential participants can be
 CC screened for the genetic basis of GS, and eliminated or included on
 CC basis of them possessing or not possessing GS.

XX Sequence 20 BP; 7 A; 6 C; 4 G; 3 T; 0 other;

Query Match 100.0%; Score 20; DB 18; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.11; Mismatches 0; Gaps 0;

QY 1 gtcacgtcacacagtcacac 20
 Db 1 gtcacgtcacacagtcacac 20

RESULT 2

AAT79540

ID AAT79540 standard; DNA; 620 BP.

XX AAT79540;

DT 23-JAN-1998 (first entry)

XX Upstream DNA sequence of UGT1*1 gene exon 1.

XX Uridine diphosphate glucuronosyltransferase gene; UGT;

KW Gilbert's syndrome; GS; unconjugated hyperbilirubinaemia;

KM bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;

XX Homo sapiens.

XX Key Location/Qualifiers

FT protein_bind 2..8

FT GC_signal 93..102

FT misc_feature 180..191

FT protein_bind 316..323

FT protein_bind 317..324

FT misc_feature 366..376

FT misc_feature 366..376

FT misc_feature 366..376

FT misc_feature 366..376

FT misc_feature 366..376

FT misc_feature 366..376

FT misc_feature 366..376

FT misc_feature 366..376

FT protein_bind 386..392

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 386..392

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 386..392

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 386..392

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 386..392

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 386..392

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

FT protein_bind 513..519

</

Db 508 gtcacgtgacacagtcacaac 527

RESULT 3

AAV13057/c
ID AAV13057 standard; DNA; 22 BP.

XX AAV13057;

XX 18-MAY-1998 (first entry)

XX 3'-flanking region of an interleukin 6 gene PCR primer.

XX Interleukin 6; IL-6; PCR primer; genotype; osteoporosis; human;
XX polymorphism; ss.

XX Synthetic.

XX Homo sapiens.

XX W09743446-A2.

XX 20-NOV-1997.

XX 16-MAY-1997; 97WO-GB01337.

XX 16-MAY-1996; 96GB-0010281.

XX (GEMI-) GEMINI INT HOLDINGS LTD.

XX Grant SFA, Raiston SH;

XX WPI; 1998-008905/01.

XX Determination of genotype of an IL-6 gene in an individual - used
XX for diagnosis of or identification of a predisposition or
XX susceptibility to a disease, particularly osteoporosis

XX Claim 9; Page 6; 20pp; English.

XX The present sequence represents a PCR primer used in the present
XX invention. The present invention describes an in vitro method of
XX diagnosis of a disease in an animal. The method comprises determining
XX the genotype of an interleukin-6 (IL-6) gene in the animal. The
XX invention also describes: (1) a method of identifying an animal
XX predisposed or susceptible to a disease, comprising determining the
XX genotype of an IL-6 gene in the animal; (2) a composition for use in
XX diagnosing a disease in an animal, the composition comprising one or
XX more primer nucleic acid molecules adapted to amplify a portion of a
XX 3' flanking region of an IL-6 gene in the animal; and (3) a composition
XX for use in identifying an animal predisposed or susceptible to a
XX disease, the composition comprising one or more primer nucleic acid
XX molecules adapted to amplify a portion of a 3' flanking region of an
XX IL-6 gene in the animal. The method can be used for the diagnosis of
XX or identifying predisposition or susceptibility to a disease,
XX particularly osteoporosis.

XX Sequence 22 BP; 4 A; 4 C; 7 G; 7 T; 0 other;

Query Match 82.0%; Score 16.4; DB 19; Length 22;
Best Local Similarity 94.4%; Pred. No. 9;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tcacgtgacacagtcacaac 19
DB 19 TCACGTGACACACTCAAA 2

RESULT 4

AAZ57627
ID AAZ57627 standard; DNA; 27 BP.

XX AAZ57627;

XX 05-APR-2000 (first entry)

XX Collagen IV alpha 3 deletion mutant fragment #1.

XX Collagen IV alpha 3; deletion mutant; nucleotide probe; genomic profile;

XX Alport syndrome; disease prognosis; health screening;

XX health management tool; ss.

XX Unidentified.

XX W0964626-A2.

XX 16-DEC-1999.

XX 04-JUN-1999; 99WO-GB01779.

XX 06-JUN-1998; 98GB-0012098.

XX 23-DEC-1998; 98GB-0028289.

XX (GENO-) GENOSTIC PHARMA LTD.

XX Roberts GW;

XX WPI; 2000-097546/08.

XX Nucleotide probes used in genetic screens for determining genomic
XX profiles, e.g. for prognosis or management -

XX Example 5; Page 25; 14pp; English.

XX This sequence is an collagen IV alpha 3 deletion mutant fragment. The
XX mutated gene is an example of a variant that can be detected using the
XX new nucleotide probes of the invention. The probes consist of
XX complementary DNA or RNA for the detection of variants in a group of
XX target genes associated with particular diseases. The mutant variant
XX represented by this sequence gives rise to Alport syndrome phenotype. The
XX specification lists approximately 2500 genes that are target groups for
XX the nucleotide probes of the invention. The probes are used to determine
XX the genomic profile of an individual, particularly for:
XX (a) prognosis and management of disease (or determining susceptibility
XX to disease);
XX (b) predicting a patient's response to therapy and symptom profiles;
XX (c) for health screening;
XX (d) to develop new strategies for therapy and clinical trials; and
XX (e) to construct health management algorithms or models, e.g. of the
XX likelihood of developing disease.
XX Antibodies specific for the proteins encoded by the genes in the target
XX list, can be used similarly, identification of a core group of genes
XX associated with disease, makes possible the use of genetic profiling as a
XX universal health management tool.

XX Sequence 27 BP; 7 A; 10 C; 3 G; 7 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 27;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcacgtgacacagtcacaac 20
DB 8 TCACGTGACACAGTCACAAC 26

RESULT 5

AAZ48039
ID AAZ48039 standard; DNA; 27 BP.

XX AAZ48039;

XX 08-MAR-2000 (first entry)

XX Human Alport syndrome oligonucleotide #1.

XX Human: glycogen storage disease 2; genetic filling; identification;
 KW polymorphic variation; mutation; detection; probe; prognosis; headache;
 KW gene association; health screening; therapy; clinical trial;
 KW absorption; distribution; metabolism; elimination; oncology; dementia;
 KW cardiovascular; gastrointestinal; respiratory; immunology; development;
 KW skin; bone; muscle; endocrine; sexual dysfunction; ss.
 XX Homo sapiens.
 OS
 PN WO9964627-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 04-JUN-1999; 99WO-GB01780.
 XX
 PR 06-JUN-1998; 98GB-0012099.
 PR 20-JUN-1998; 98GB-0013291.
 PR 24-JUN-1998; 98GB-0013611.
 PR 27-JUN-1998; 98GB-0013835.
 PR 01-JUL-1998; 98GB-0014110.
 PR 07-JUL-1998; 98GB-0014580.
 PR 16-JUL-1998; 98GB-0015438.
 PR 18-JUL-1998; 98GB-0015574.
 PR 18-JUL-1998; 98GB-0015576.
 PR 24-JUL-1998; 98GB-0016085.
 PR 24-JUL-1998; 98GB-0016086.
 PR 05-AUG-1998; 98GB-0016921.
 PR 07-AUG-1998; 98GB-0017097.
 PR 08-AUG-1998; 98GB-0017200.
 PR 14-AUG-1998; 98GB-0017632.
 PR 19-AUG-1998; 98GB-0017943.
 XX
 XX (GENO-) GENOSTIC PHARMA LTD.
 XX
 PI Roberts GM;
 XX
 DR MPI: 2000-097547/08.
 XX
 PT Nucleotide probes used in genetic screens for determining genomic
 profiles, e.g. for prognosis or management -
 PS Example 5; Page 25; 745pp; English.
 XX
 CC The present invention describes new nucleotide probes (i) comprising
 CC complementary DNA or RNA, used in the detection of variants in a group
 CC of target genes associated with particular diseases. Detected variants
 CC are mutations or polymorphisms. Included in the specification are lists
 CC of genes for the following target groups: (i) ADME (absorption,
 CC distribution, metabolism and elimination - about 1000 genes);
 CC (ii) oncology (about 1200 genes); (iii) central nervous system (about
 CC 1100 genes); (iv) dementia (about 250 genes); (v) brain injury (about
 CC 850 genes); (vi) behavioural (about 250 genes); (vii) psychoses and
 CC personality (about 500 genes); (viii) cardiovascular (about 1100 genes);
 CC (ix) gastrointestinal (about 1100 genes); (x) respiratory (about 800
 CC genes); (xi) immunology (about 1100 genes); (xii) development (about 2000
 CC and endocrine (about 1100 genes); (xiii) headache (about 250 genes); and
 CC (xiv) sexual dysfunction (about 250 genes); (xv) metabolism
 CC genomic profile of an individual, particularly for: (a) prognosis and
 CC management of disease (or determining susceptibility to disease);
 CC (b) predicting a patient's response to therapy and symptom profiles;
 CC (c) for health screening; (d) to develop new strategies for therapy and
 CC clinical trials; and (e) to construct health management algorithms or
 CC models, e.g. of the likelihood of developing disease. The present
 CC sequence represents an oligonucleotide used to illustrate a deletion in
 CC Alport syndrome, which is given in an example from the present invention.
 XX
 SQ Sequence 27 BP; 7 A; 10 C; 3 G; 7 T; 0 other;

Query Match

79.0%; Score 15.8; DB 21; Length 27;

Best Local Similarity 89.5%; Pred. No. 19;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 tcacgtgcacacgtcaaac 20
 Db ||||| ||||| |||||
 8 tcaccgcacacgtcaaac 26

RESULT 6
 AAX98515/C
 ID AAX98515 standard; cDNA: 300 BP.
 AC AAX98515;
 XX
 DT 24-SEP-1999 (first entry)
 XX
 DE Human cancer cell derived cDNA #241.
 XX
 KW Cancer; human; colon; breast; lung; transmembrane receptor; ATPase;
 KW integral membrane protein; aspartyl protease; GATA family; wt family;
 KW transcription factor; G-protein alpha subunit; protein phosphatase;
 KW photolipase binding protein; diacylglycerol binding protein; trypsin;
 KW protein kinase; tyrosine phosphatase; developmental signalling protein;
 KW Wnt/RSF/Wnt domain; therapy; forensic; genetic mapping; diagnostic;
 KW detection; treatment; cervical; melanoma; colorectal adenocarcinoma;
 KW Wilms' tumour; retinoblastoma; sarcoma; myosarcoma; lung carcinoma;
 KW leukemia; lymphoma; dysplasia; hyperplasia; endometrium; adrenal;
 KW prostate; ss.
 KW
 KW
 OS Homo sapiens.
 XX
 XX WO9933982-A2.
 XX
 XX 08-JUL-1999.
 XX
 XX
 XX 22-DEC-1998; 98WO-US27610.
 XX
 XX 21-DEC-1998; 98US-0217471.
 XX 23-DEC-1997; 97US-0068755.
 XX 03-APR-1998; 98US-0080664.
 XX 21-OCT-1998; 98US-0105234.
 XX 27-OCT-1998; 98US-0105877.
 XX
 PA (CHIR) CHIRON CORP.
 XX (HYSE-) HXSEQ INC.
 XX
 PI Crivenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones LW, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lanson G, Leshkowitz D, Pol D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 DR MPI: 1999-430243/36.
 XX
 XX New isolated human polynucleotides
 PS Claim 1; Page 375; 591pp; English.
 XX
 CC This invention describes novel isolated human polynucleotides obtained
 CC by screening for differential expression in colon cancer, breast cancer
 CC and lung cancer cell lines. The polynucleotides of the invention are
 CC represented in AAX98275-X9918 and encode polypeptides of protein
 CC families selected from 4 transmembrane segments integral membrane
 CC proteins, 7 transmembrane segments integral membrane
 CC cellular activities (AA), eukaryotic aspartyl proteases, GATA family of
 CC transcription factors (AA), eukaryotic aspartyl proteases, GATA family of
 CC diacylglycerol binding proteins, protein kinase, protein phosphatase 2C,
 CC signalling proteins and Wnt/RSF/Wnt domain containing proteins. The
 CC encoded polypeptides also have a functional domain selected from Ank
 CC bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger
 CC (C2H2 type), zinc finger (C2HC class), and zinc-binding metalloprotease

CC domain. The polynucleotides encode polypeptides with similarity to known
 CC protein families and are predicted to have similar properties. The novel
 CC polynucleotides can be used to develop products for use as therapeutic
 CC agents and in forensics, genetic analysis, mapping and diagnostic
 CC applications. In particular, the product can be used for the detection
 CC and management of cancers. They can be used for treating e.g. cervical
 CC cancers, melanomas, colorectal adenocarcinomas, Wilms' tumour, sarcomas,
 CC retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic
 CC myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and
 CC myeloid leukemia, and lymphomas such as histiocytic lymphoma, anaplastic
 CC histiocytic sarcoma, and lymphomas of the cervix, fibrous dysplasia of bone, and
 CC epithelial dysplasia of the cervix, fibrous dysplasia of bone, and
 CC mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast,
 CC prostate or thyroid hyperplasias or pseudoeptitheliomatous hyperplasia of
 CC the skin.

CC Sequence 300 BP; 65 A; 110 C; 40 G; 85 T; 0 other;

Query Match 79.0%; Score 15.8; DB 20; Length 300;
 Best Local Similarity 89.5%; Pred. No. 28;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtccgtgacacagtcacaa 19
 |||||
 DB 206 gtccgtgacacagtcacaa 188

RESULT 7
 ID AAA50369 standard; DNA; 392 BP.

AC AAA50369;

XX 20-NOV-2000 (first entry)

DE Human Goodpasture antigen DeltaIII/IV/ V cDNA.

KW Goodpasture antigen; GPdeltaIII/IV/ V; human; GPP;

KM goodpasture antigen binding protein; autoimmune disease; apoptosis;

KK cancer; tumour; gene therapy; ss.

XX Homo sapiens.

OS Key Location/Qualifiers

FT CDS 1..210
 FT /*tag= a
 FT /product= Met-deleted GPdeltaIII/IV/ V
 FT /partial

PN W0200050607-A2.

XX 31-AUG-2000.

PD 24-FEB-2000; 2000MO-IB00324.

PF 24-FEB-1999; 99US-0121483.

PR (SAUS/) SAUS J.

XX Saus J;

XX WPI; 2000-572094/53.

DR P-PSDB; AAY95920.

XX Novel Goodpasture antigen binding proteins useful for diagnosing and
 XX treating autoimmune disorders, tumor, and preventing cell apoptosis -

XX Claim 25; Page 153-156; 156pp; English.

CC The present sequence is that of cDNA encoding human Goodpasture
 CC antigen (GP) DeltaIII/IV/ V (see AAY95920), an alternative form of
 CC human GP resulting from splicing out of exons III, IV and V. The

CC cDNA was obtained by subcloning a cDNA encoding the protein into a
 CC modified pET5b vector including an initiator Met. The invention
 CC relates to novel Goodpasture antigen binding proteins (GPPs, see
 CC AAY95900-11), which bind to and phosphorylate the unique N-terminal
 CC region of human GP, and which are highly expressed in several
 CC autoimmune conditions. Claimed methods for treating an autoimmune
 CC disorder, cell apoptosis or a tumour involve modifying the
 CC expression or activity of GPP, especially using a GP-derived
 CC peptide, such as GPdeltaIII/IV/ V or a nucleic acid sequence
 CC encoding it.

XX Sequence 392 BP; 134 A; 80 C; 81 G; 97 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 392;
 Best Local Similarity 89.5%; Pred. No. 29;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcacgtgacacagtcacaa 20
 |||||
 DB 65 tcacgtgacacagtcacaa 83

RESULT 8
 ID AAA50370 standard; DNA; 507 BP.

AC AAA50370;

XX 20-NOV-2000 (first entry)

DE Human Goodpasture antigen DeltaIII/ V cDNA.

KW Goodpasture antigen; GPdeltaIII/IV/ V; human; GPP;

KM goodpasture antigen binding protein; autoimmune disease; apoptosis;

KK cancer; tumour; gene therapy; ss.

XX Homo sapiens.

OS Key Location/Qualifiers

FT CDS 1..219
 FT /*tag= a
 FT /product= Met-deleted GPdeltaIII/IV/ V
 FT /partial

PN W0200050607-A2.

XX 31-AUG-2000.

PD 24-FEB-2000; 2000MO-IB00324.

PF 24-FEB-1999; 99US-0121483.

PR (SAUS/) SAUS J.

XX Saus J;

XX WPI; 2000-572094/53.

DR P-PSDB; AAY95921.

XX Novel Goodpasture antigen binding proteins useful for diagnosing and
 XX treating autoimmune disorders, tumor, and preventing cell apoptosis -

XX Claim 25; Page 154-155; 156pp; English.

CC The present sequence is that of cDNA encoding human Goodpasture
 CC antigen (GP) DeltaIII/IV/ V (see AAY95921), an alternative form of
 CC human GP resulting from splicing out of exons III and V. The
 CC cDNA was obtained by subcloning a cDNA encoding the protein into a
 CC modified pET5b vector including an initiator Met. The invention
 CC relates to novel Goodpasture antigen binding proteins (GPPs, see
 CC AAY95900-11), which bind to and phosphorylate the unique N-terminal
 CC region of human GP, and which are highly expressed in several

CC autoimmune conditions. Claimed methods for treating an autoimmune
 CC disorder, cell apoptosis or a tumour involve modifying the
 CC expression or activity of GPBP, especially using a GP-derived
 CC peptide, such as GPdeltaII/V or a nucleic acid sequence encoding
 CC it.

XX Sequence 507 BP; 159 A; 113 C; 104 G; 131 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 507;
 Best Local Similarity 89.5%; Pred. No. 31;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcacgtgacacagtcacaac 20
 ||||| ||||| ||||| |||||
 Db 65 tcaccgcacacagtcacaac 83

RESULT 9

AAAS0368 standard; DNA; 680 BP.

AC AAAS0368;

DT 20-NOV-2000 (first entry)

DE Human Goodpasture antigen DeltaIII cDNA.

KW Goodpasture antigen; GPdeltaIII; human; GPBP;

KW Goodpasture antigen binding protein; autoimmune disease; apoptosis;
 cancer; tumour; gene therapy; ss.

OS Homo sapiens.

FT Key

FT CDS

Location/Qualifiers
 1..219
 /*tag- a
 /product- Met-deleted GPdeltaIII
 /partial

MO200050607-A2.

PD 31-AUG-2000.

PF 24-FEB-2000; 2000MO-IB00324.

PR 24-FEB-1999; 99US-0121483.

PA (SAUS/) SAUS J.

PI Saus J;

DR WPI: 2000-572094/53.

DR P-PSDB: AAY95919.

PT Novel Goodpasture antigen binding proteins useful for diagnosing and
 PT treating autoimmune disorders, tumor, and preventing cell apoptosis
 PS Claim 25; Page 152-153; 158pp; English.

CC The present sequence is that of cDNA encoding human Goodpasture
 CC antigen (GP) DeltaIII (see AAY95919), i.e. an alternative form of
 CC human GP resulting from splicing out of exon III. The cDNA was
 CC obtained by subcloning a cDNA encoding the protein into a modified
 CC pET15b vector including an initiator Met. The invention relates
 CC to novel Goodpasture antigen binding proteins (GPBs), see
 CC AAY95900-11), which bind to and phosphorylate the unique N-terminal
 CC region of human GP, and which are highly expressed in several
 CC autoimmune conditions. Claimed methods for treating an autoimmune
 CC disorder, cell apoptosis or a tumour involve modifying the
 CC expression or activity of GPBP, especially using a GP-derived
 CC peptide, such as GPdeltaIII or a nucleic acid sequence encoding it.

SO Sequence 680 BP; 204 A; 159 C; 145 G; 172 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 680;
 Best Local Similarity 89.5%; Pred. No. 32;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcacgtgacacagtcacaac 20
 ||||| ||||| ||||| |||||
 Db 65 tcaccgcacacagtcacaac 83

RESULT 10

AAAS0367 standard; DNA; 685 BP.

AC AAAS0367;

DT 20-NOV-2000 (first entry)

DE Human Goodpasture antigen DeltaV cDNA.

KW Goodpasture antigen; GPdeltaV; goodpasture antigen binding protein;

KW GPBP; human; autoimmune disease; apoptosis; cancer; tumour;
 gene therapy; ss.

OS Homo sapiens.

FT Key

FT CDS

Location/Qualifiers
 1..636
 /*tag- a
 /product- Met-deleted GPdeltaV
 /partial

MO200050607-A2.

PD 31-AUG-2000.

PF 24-FEB-2000; 2000MO-IB00324.

PR 24-FEB-1999; 99US-0121483.

PA (SAUS/) SAUS J.

PI Saus J;

DR WPI: 2000-572094/53.

DR P-PSDB: AAY95918.

PT Novel Goodpasture antigen binding proteins useful for diagnosing and
 PT treating autoimmune disorders, tumor, and preventing cell apoptosis
 PS Claim 25; Page 150-151; 158pp; English.

CC The present sequence is that of cDNA encoding human Goodpasture
 CC antigen (GP) DeltaV (see AAY95918), i.e. an alternative form of
 CC human GP resulting from splicing out of exon V. The cDNA was
 CC obtained by subcloning a cDNA encoding the protein into a modified
 CC pET15b vector including an initiator Met. The invention relates
 CC to novel Goodpasture antigen binding proteins (GPBs), see
 CC AAY95900-11), which bind to and phosphorylate the unique N-terminal
 CC region of human GP, and which are highly expressed in several
 CC autoimmune conditions. Claimed methods for treating an autoimmune
 CC disorder, cell apoptosis or a tumour involve modifying the
 CC expression or activity of GPBP, especially using a GP-derived
 CC peptide, such as GPdeltaV or a nucleic acid sequence encoding it.

SO Sequence 685 BP; 206 A; 157 C; 138 G; 184 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 685;
 Best Local Similarity 89.5%; Pred. No. 32;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2 tcaagctgacacagtcacac 20

Disclosure: Fig 17c; 56pp; Engl15h.

2 taacgtgacacgtcaaac 20

Enlighten

XX This is the nucleotide sequence of a recombinant DNA encoding a
 CC type IV collagen non-collagenous (NC1) domain alpha-3 polypeptide
 CC (see AAY1393) composed of a BM40 signal sequence (which is cleaved
 CC from the mature protein) to facilitate protein secretion, and a
 CC mature protein comprising an affinity tag (facilitates purification
 CC and identification of the material) and the alpha-1 chain monomer.
 CC The invention provides methods and kits for inhibiting angiogenesis,
 CC tumour growth and metastasis, and endothelial cell interaction with
 CC the extracellular matrix, each method comprising contacting the
 CC tumour or animal tissue with 1 or more isolated type IV collagen
 CC NC1 alpha chain monomer(s) selected from the group consisting of
 CC alpha-1, alpha-2, alpha-3 and alpha-6 NC1 chain monomers. The
 CC monomers can be produced via recombinant protein expression. The
 CC polynucleotides and polypeptides are used to treat an angiogenesis-
 CC mediated disorder or condition, especially selected from solid and
 CC blood-borne tumours, diabetic retinopathy, rheumatoid arthritis,
 CC retinal neovascularization, choroidal neovascularization, macular
 CC degeneration, corneal neovascularization, retinopathy of prematurity,
 CC corneal graft rejection, neovascular glaucoma, retrolental
 CC fibroplasia, epidemic keratoconjunctivitis, vitamin A deficiency,
 CC contact lens overwear, atopic keratitis, superior limbic keratitis,
 CC pterygium keratitis sicca, seborrheic acne rosacea, phlyctenulosis,
 CC syphilis, mycobacteria infections, lipid degeneration, chemical
 CC burns, bacterial ulcers, fungal ulcers, herpes simplex infections,
 CC herpes zoster infections, protozoan infections, Kaposi's sarcoma,
 CC Mooren ulcer, Terrien's marginal degeneration, marginal keratolysis,
 CC keratitis, Steven's Johnson disease, radial keratotomy, sickle cell
 CC anemia, sarcoid, pseudoxanthoma elasticum, Paget's disease, vein
 CC occlusion, artery occlusion, carotid obstructive disease, chronic
 CC uveitis, chronic vitritis, Lyme's disease, Bala's disease, Behcet's
 CC disease, myopia, optic pits, Stargardt's disease, pars planitis,
 CC chronic retinal detachment, hyperviscosity syndromes, toxoplasmosis,
 CC post-laser complications, abnormal proliferation of fibrovascular
 CC tissue, haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular
 CC disease, osteoarthritis, chronic inflammation, Crohn's disease,
 CC ulcerative colitis, psoriasis, atherosclerosis, and pemphigoid (all
 CC claimed).

SQ Sequence 900 BP; 228 A; 243 C; 206 G; 223 T; 0 other;

Query Match 79.0%; Score 15.8; DB 20; Length 900;
 Best Local Similarity 89.5%; Pred. No. 34;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcacgtgacacagtcacaac 20
 |||| |||||
 Db 176 tcaccgcgacacagtcacaac 194

RESULT 13
 AAA90993
 ID AAA90993 standard; DNA; 900 BP.

AC AAA90993;

DT 12-FEB-2001 (first entry)

DE Human alpha3(IV)NC1 coding sequence.

XX Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;
 KW tumour growth; integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;
 KW retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukaemia;
 KW diabetic retinopathy; rheumatoid arthritis; neovascularisation;
 KW muscular degeneration; corneal graft rejection; vitamin A deficiency;
 KW atopic keratitis; Mycobacteria infection; chemical burn; sarcoid;
 KW Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;
 KW chronic inflammation; psoriasis; therapy; alpha3(IV)NC1; ds.

OS Homo sapiens.

XX

PH Key Location/Qualifiers
 FT COS 40..846
 FT /*tag= a
 FT /product= alpha3(IV)NC1

XX WO200059532-A1.
 XX 12-OCT-2000.
 XX 31-MAR-2000; 2000WO-0508678.
 XX 01-APR-1999; 99US-0127391.
 XX (BIOS-) BIOSTRATUM INC.
 XX Brooks P, Hudson B;
 XX WPI; 2000-664962/64.
 XX P-PDB; AAY97555.

PT Use of antagonists of specific integrin receptors for inhibiting
 PT angiogenesis, tumour growth or metastases, or endothelial cell
 PT interactions with the extracellular matrix

PS Disclosures: Flg 17c; 78pp; English.

CC This sequence encodes a human type IV collagen alpha chain monomer,
 CC designated alpha3(IV)NC1. The invention relates to a method for
 CC inhibiting angiogenesis, tumour growth or metastases, or endothelial cell
 CC interactions with the extracellular matrix, comprising contacting the
 CC cells or tissue with a polypeptide composition containing antagonists of
 CC specific integrin receptors. The methods and the antagonists are useful
 CC for inhibiting angiogenesis, tumour growth or metastases, or endothelial
 CC cell interaction with the extracellular matrix. The antagonists are also
 CC useful for treating diseases and conditions with accompanying undesired
 CC angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,
 CC carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,
 CC neuroblastoma, osteosarcoma or leukaemia). These are also applicable to
 CC treating non-tumorigenic diseases and conditions with accompanying
 CC undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis,
 CC retinal neovascularisation, choroidal neovascularisation, muscular
 CC degeneration, corneal graft rejection, vitamin A deficiency, atopic
 CC keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,
 CC sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser
 CC complications, chronic inflammation or psoriasis.

SQ Sequence 900 BP; 228 A; 243 C; 206 G; 223 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 900;
 Best Local Similarity 89.5%; Pred. No. 34;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcacgtgacacagtcacaac 20
 |||| |||||
 Db 176 tcaccgcgacacagtcacaac 194

RESULT 14

AAFL13921/C

ID AAFL13921 standard; cDNA; 1276 BP.

AC AAFL13921;

DT 13-MAR-2001 (first entry)

DE Aspergillus oryzae EST SEQ ID NO:6444.

XX Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX OS Aspergillus oryzae.
XX PN WO200056762-A2.
XX PD 28-SEP-2000.
XX PF 22-MAR-2000; 2000WO-US07781.
XX PR 22-MAR-1999; 99US-0273623.
XX PA (NOVO) NOVO NORDISK BIOTECH INC.
XX PA (NOVO) NOVO NORDISK AS.
XX PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX DR WPI: 2000-594572/56.
XX PT Monitoring differential expression of genes in filamentous fungal cells
XX PT uses fluorescence-labeled nucleic acids isolated from the cells and a
XX PT substrate of expressed sequence tags -
XX PS Claim 88: Page 2643-2644; 3161pp; English.
XX CC The present invention describes a method for monitoring differential
XX CC expression of genes in a first filamentous fungal (FF) cell relative to
XX CC expression of the same genes in one or more second filamentous fungal
XX CC cells. The method uses fluorescence-labeled nucleic acids isolated from
XX CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX CC are used in the methods for monitoring differential expression of genes
XX CC in a first filamentous fungal (FF) cell relative to expression of the
XX CC same genes in one or more second filamentous fungal cells. Monitoring
XX CC the global expression of genes from FF cells allows the production
XX CC potential of the microorganisms to be improved. New genes may be
XX CC discovered, possible functions of unknown open reading frames can be
XX CC identified and gene copy number variation and stability can be
XX CC monitored. The expression of genes can be used to study how FF cells
XX CC adapt to changes in culture conditions, environmental stress, spore
XX CC morphogenesis, recombination, metabolic or catabolic pathway
XX CC engineering. Using ESTs provides several advantages over genomic or
XX CC random cDNA clones including elimination of redundancy as one spot on an
XX CC array equals one gene or open reading frame, and organization of the
XX CC microarrays based on function of the gene products to facilitate
XX CC analysis of the results. AA070478 to AA071247 represents ESTs from
XX CC Fusarium venenatum; AA071248 to AA071833 represents ESTs from Aspergillus
XX CC niger; AA071834 to AA071878 represents ESTs from Aspergillus oryzae; and
XX CC AA071879 to AA071937 represents ESTs from Trichoderma reesei, which are
XX CC all specifically claimed in the present invention.
XX SS Sequence 1276 BP; 311 A; 319 C; 314 G; 332 T; 0 other;

Query Match 76.0%; Score 15.2; DB 21; Length 1276;
Best Local Similarity 85.0%; Pred. NO. 73;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

CY 1 gtcacgtacacagctaacac 20
DB 746 gtccttgacacgacgacac 727

RESULT 15
AAC35997
ID AAC35997 standard; DNM; 1801 BP.
XX AAC35997;
XX AC AAC35997;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SMO ID NO: 12116.
XX AA Hybridisation assay; genetic mapping; gene expression control;
XX KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX OS EP1033405-A2.
XX PN 06-SEP-2000.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0133456.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134421.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
XX PR 18-JUN-1999; 99US-0139462.
XX PR 18-JUN-1999; 99US-0139463.
XX PR 18-JUN-1999; 99US-0139464.
XX PR 18-JUN-1999; 99US-0139465.
XX PR 18-JUN-1999; 99US-0139466.
XX PR 18-JUN-1999; 99US-0139467.
XX PR 18-JUN-1999; 99US-0139468.
XX PR 18-JUN-1999; 99US-0139469.
XX PR 18-JUN-1999; 99US-0139470.
XX PR 18-JUN-1999; 99US-0139471.
XX PR 18-JUN-1999; 99US-0139472.
XX PR 18-JUN-1999; 99US-0139473.
XX PR 18-JUN-1999; 99US-0139474.
XX PR 18-JUN-1999; 99US-0139475.
XX PR 18-JUN-1999; 99US-0139476.
XX PR 18-JUN-1999; 99US-0139477.
XX PR 18-JUN-1999; 99US-0139478.
XX PR 18-JUN-1999; 99US-0139479.
XX PR 18-JUN-1999; 99US-0139480.
XX PR 18-JUN-1999; 99US-0139481.
XX PR 18-JUN-1999; 99US-0139482.
XX PR 18-JUN-1999; 99US-0139483.
XX PR 18-JUN-1999; 99US-0139484.
XX PR 18-JUN-1999; 99US-0139485.
XX PR 18-JUN-1999; 99US-0139486.
XX PR 18-JUN-1999; 99US-0139487.
XX PR 18-JUN-1999; 99US-0139488.
XX PR 18-JUN-1999; 99US-0139489.
XX PR 18-JUN-1999; 99US-0139490.
XX PR 18-JUN-1999; 99US-0139491.
XX PR 18-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139493.
XX PR 18-JUN-1999; 99US-0139494.
XX PR 18-JUN-1999; 99US-0139495.
XX PR 18-JUN-1999; 99US-0139496.
XX PR 18-JUN-1999; 99US-0139497.
XX PR 18-JUN-1999; 99US-0139498.
XX PR 18-JUN-1999; 99US-0139499.
XX PR 18-JUN-1999; 99US-0139500.
XX PR 18-JUN-1999; 99US-0139501.
XX PR 18-JUN-1999; 99US-0139502.
XX PR 18-JUN-1999; 99US-0139503.
XX PR 18-JUN-1999; 99US-0139504.
XX PR 18-JUN-1999; 99US-0139505.
XX PR 18-JUN-1999; 99US-0139506.
XX PR 18-JUN-1999; 99US-0139507.
XX PR 18-JUN-1999; 99US-0139508.
XX PR 18-JUN-1999; 99US-0139509.
XX PR 18-JUN-1999; 99US-0139510.
XX PR 18-JUN-1999; 99US-0139511.
XX PR 18-JUN-1999; 99US-0139512.
XX PR 18-JUN-1999; 99US-0139513.
XX PR 18-JUN-1999; 99US-0139514.
XX PR 18-JUN-1999; 99US-0139515.
XX PR 18-JUN-1999; 99US-0139516.
XX PR 18-JUN-1999; 99US-0139517.
XX PR 18-JUN-1999; 99US-0139518.
XX PR 18-JUN-1999; 99US-0139519.
XX PR 18-JUN-1999; 99US-0139520.
XX PR 18-JUN-1999; 99US-0139521.
XX PR 18-JUN-1999; 99US-0139522.
XX PR 18-JUN-1999; 99US-0139523.
XX PR 18-JUN-1999; 99US-0139524.
XX PR 18-JUN-1999; 99US-0139525.
XX PR 18-JUN-1999; 99US-0139526.
XX PR 18-JUN-1999; 99US-0139527.
XX PR 18-JUN-1999; 99US-0139528.
XX PR 18-JUN-1999; 99US-0139529.
XX PR 18-JUN-1999; 99US-0139530.
XX PR 18-JUN-1999; 99US-0139531.
XX PR 18-JUN-1999; 99US-0139532.
XX PR 18-JUN-1999; 99US-0139533.
XX PR 18-JUN-1999; 99US-0139534.
XX PR 18-JUN-1999; 99US-0139535.
XX PR 18-JUN-1999; 99US-0139536.
XX PR 18-JUN-1999; 99US-0139537.
XX PR 18-JUN-1999; 99US-0139538.
XX PR 18-JUN-1999; 99US-0139539.
XX PR 18-JUN-1999; 99US-0139540.
XX PR 18-JUN-1999; 99US-0139541.
XX PR 18-JUN-1999; 99US-0139542.
XX PR 18-JUN-1999; 99US-0139543.
XX PR 18-JUN-1999; 99US-0139544.
XX PR 18-JUN-1999; 99US-0139545.
XX PR 18-JUN-1999; 99US-0139546.
XX PR 18-JUN-1999; 99US-0139547.
XX PR 18-JUN-1999; 99US-0139548.
XX PR 18-JUN-1999; 99US-0139549.
XX PR 18-JUN-1999; 99US-0139550.
XX PR 18-JUN-1999; 99US-0139551.
XX PR 18-JUN-1999; 99US-0139552.
XX PR 18-JUN-1999; 99US-0139553.
XX PR 18-JUN-1999; 99US-0139554.
XX PR 18-JUN-1999; 99US-0139555.
XX PR 18-JUN-1999; 99US-0139556.
XX PR 18-JUN-1999; 99US-0139557.
XX PR 18-JUN-1999; 99US-0139558.
XX PR 18-JUN-1999; 99US-0139559.
XX PR 18-JUN-1999; 99US-0139560.
XX PR 18-JUN-1999; 99US-0139561.
XX PR 18-JUN-1999; 99US-0139562.
XX PR 18-JUN-1999; 99US-0139563.
XX PR 18-JUN-1999; 99US-0139564.
XX PR 18-JUN-1999; 99US-0139565.
XX PR 18-JUN-1999; 99US-0139566.
XX PR 18-JUN-1999; 99US-0139567.
XX PR 18-JUN-1999; 99US-0139568.
XX PR 18-JUN-1999; 99US-0139569.
XX PR 18-JUN-1999; 99US-0139570.
XX PR 18-JUN-1999; 99US-0139571.
XX PR 18-JUN-1999; 99US-0139572.
XX PR 18-JUN-1999; 99US-0139573.
XX PR 18-JUN-1999; 99US-0139574.
XX PR 18-JUN-1999; 99US-0139575.
XX PR 18-JUN-1999; 99US-0139576.
XX PR 18-JUN-1999; 99US-0139577.
XX PR 18-JUN-1999; 99US-0139578.
XX PR 18-JUN-1999; 99US-0139579.
XX PR 18-JUN-1999; 99US-0139580.
XX PR 18-JUN-1999; 99US-0139581.
XX PR 18-JUN-1999; 99US-0139582.
XX PR 18-JUN-1999; 99US-0139583.
XX PR 18-JUN-1999; 99US-0139584.
XX PR 18-JUN-1999; 99US-0139585.
XX PR 18-JUN-1999; 99US-0139586.
XX PR 18-JUN-1999; 99US-0139587.
XX PR 18-JUN-1999; 99US-0139588.
XX PR 18-JUN-1999; 99US-0139589.
XX PR 18-JUN-1999; 99US-0139590.
XX PR 18-JUN-1999; 99US-0139591.
XX PR 18-JUN-1999; 99US-0139592.
XX PR 18-JUN-1999; 99US-0139593.
XX PR 18-JUN-1999; 99US-0139594.
XX PR 18-JUN-1999; 99US-0139595.
XX PR 18-JUN-1999; 99US-0139596.
XX PR 18-JUN-1999; 99US-0139597.
XX PR 18-JUN-1999; 99US-0139598.
XX PR 18-JUN-1999; 99US-0139599.
XX PR 18-JUN-1999; 99US-0139600.
XX PR 18-JUN-1999; 99US-0139601.
XX PR 18-JUN-1999; 99US-0139602.
XX PR 18-JUN-1999; 99US-0139603.
XX PR 18-JUN-1999; 99US-0139604.
XX PR 18-JUN-1999; 99US-0139605.
XX PR 18-JUN-1999; 99US-0139606.
XX PR 18-JUN-1999; 99US-0139607.
XX PR 18-JUN-1999; 99US-0139608.
XX PR 18-JUN-1999; 99US-0139609.
XX PR 18-JUN-1999; 99US-0139610.
XX PR 18-JUN-1999; 99US-0139611.
XX PR 18-JUN-1999; 99US-0139612.
XX PR 18-JUN-1999; 99US-0139613.
XX PR 18-JUN-1999; 99US-0139614.
XX PR 18-JUN-1999; 99US-0139615.
XX PR 18-JUN-1999; 99US-0139616.
XX PR 18-JUN-1999; 99US-0139617.
XX PR 18-JUN-1999; 99US-0139618.
XX PR 18-JUN-1999; 99US-0139619.
XX PR 18-JUN-1999; 99US-0139620.
XX PR 18-JUN-1999; 99US-0139621.
XX PR 18-JUN-1999; 99US-0139622.
XX PR 18-JUN-1999; 99US-0139623.
XX PR 18-JUN-1999; 99US-0139624.
XX PR 18-JUN-1999; 99US-0139625.
XX PR 18-JUN-1999; 99US-0139626.
XX PR 18-JUN-1999; 99US-0139627.
XX PR 18-JUN-1999; 99US-0139628.
XX PR 18-JUN-1999; 99US-0139629.
XX PR 18-JUN-1999; 99US-0139630.
XX PR 18-JUN-1999; 99US-0139631.
XX PR 18-JUN-1999; 99US-0139632.
XX PR 18-JUN-1999; 99US-0139633.
XX PR 18-JUN-1999; 99US-0139634.
XX PR 18-JUN-1999; 99US-0139635.
XX PR 18-JUN-1999; 99US-0139636.
XX PR 18-JUN-1999; 99US-0139637.
XX PR 18-JUN-1999; 99US-0139638.
XX PR 18-JUN-1999; 99US-0139639.
XX PR 18-JUN-1999; 99US-0139640.
XX PR 18-JUN-1999; 99US-0139641.
XX PR 18-JUN-1999; 99US-0139642.
XX PR 18-JUN-1999; 99US-0139643.
XX PR 18-JUN-1999; 99US-0139644.
XX PR 18-JUN-1999; 99US-0139645.
XX PR 18-JUN-1999; 99US-0139646.
XX PR 18-JUN-1999; 99US-0139647.
XX PR 18-JUN-1999; 99US-0139648.
XX PR 18-JUN-1999; 99US-0139649.
XX PR 18-JUN-1999; 99US-0139650.
XX PR 18-JUN-1999; 99US-0139651.
XX PR 18-JUN-1999; 99US-0139652.
XX PR 18-JUN-1999; 99US-0139653.
XX PR 18-JUN-1999; 99US-0139654.
XX PR 18-JUN-1999; 99US-0139655.
XX PR 18-JUN-1999; 99US-0139656.
XX PR 18-JUN-1999; 99US-0139657.
XX PR 18-JUN-1999; 99US-0139658.
XX PR 18-JUN-1999; 99US-0139659.
XX PR 18-JUN-1999; 99US-0139660.
XX PR 18-JUN-1999; 99US-0139661.
XX PR 18-JUN-1999; 99US-0139662.
XX PR 18-JUN-1999; 99US-0139663.
XX PR 18-JUN-1999; 99US-0139664.
XX PR 18-JUN-1999; 99US-0139665.
XX PR 18-JUN-1999; 99US-0139666.
XX PR 18-JUN-1999; 99US-0139667.
XX PR 18-JUN-1999; 99US-0139668.
XX PR 18-JUN-1999; 99US-0139669.
XX PR 18-JUN-1999; 99US-0139670.
XX PR 18-JUN-1999; 99US-0139671.
XX PR 18-JUN-1999; 99US-0139672.
XX PR 18-JUN-1999; 99US-0139673.
XX PR 18-JUN-1999; 99US-0139674.
XX PR 18-JUN-1999; 99US-0139675.
XX PR 18-JUN-1999; 99US-0139676.
XX PR 18-JUN-1999; 99US-0139677.
XX PR 18-JUN-1999; 99US-0139678.
XX PR 18-JUN-1999; 99US-0139679.
XX PR 18-JUN-1999; 99US-0139680.
XX PR 18-JUN-1999; 99US-0139681.
XX PR 18-JUN-1999; 99US-0139682.
XX PR 18-JUN-1999; 99US-0139683.
XX PR 18-JUN-1999; 99US-0139684.
XX PR 18-JUN-1999; 99US-0139685.
XX PR 18-JUN-1999; 99US-0139686.
XX PR 18-JUN-1999; 99US-0139687.
XX PR 18-JUN-1999; 99US-0139688.
XX PR 18-JUN-1999; 99US-0139689.
XX PR 18-JUN-1999; 99US-0139690.
XX PR 18-JUN-1999; 99US-0139691.
XX PR 18-JUN-1999; 99US-0139692.
XX PR 18-JUN-1999; 99US-0139693.
XX PR 18-JUN-1999; 99US-0139694.
XX PR 18-JUN-1999; 99US-0139695.
XX PR 18-JUN-1999; 99US-0139696.
XX PR 18-JUN-1999; 99US-0139697.
XX PR 18-JUN-1999; 99US-0139698.
XX PR 18-JUN-1999; 99US-0139699.
XX PR 18-JUN-1999; 99US-0139700.
XX PR 18-JUN-1999; 99US-0139701.
XX PR 18-JUN-1999; 99US-0139702.
XX PR 18-JUN-1999; 99US-0139703.
XX PR 18-JUN-1999; 99US-0139704.
XX PR 18-JUN-1999; 99US-0139705.
XX PR 18-JUN-1999; 99US-0139706.
XX PR 18-JUN-1999; 99US-0139707.
XX PR 18-JUN-1999; 99US-0139708.
XX PR 18-JUN-1999; 99US-0139709.
XX PR 18-JUN-1999; 99US-0139710.
XX PR 18-JUN-1999; 99US-0139711.
XX PR 18-JUN-1999; 99US-0139712.
XX PR 18-JUN-1999; 99US-0139713.
XX PR 18-JUN-1999; 99US-0139714.
XX PR 18-JUN-1999; 99US-0139715.
XX PR 18-JUN-1999; 99US-0139716.
XX PR 18-JUN-1999; 99US-0139717.
XX PR 18-JUN-1999; 99US-0139718.
XX PR 18-JUN-1999; 99US-0139719.
XX PR 18-JUN-1999; 99US-0139720.
XX PR 18-JUN-1999; 99US-0139721.
XX PR 18-JUN-1999; 99US-0139722.
XX PR 18-JUN-1999; 99US-0139723.
XX PR 18-JUN-1999; 99US-0139724.
XX PR 18-JUN-1999; 99US-0139725.
XX PR 18-JUN-1999; 99US-0139726.
XX PR 18-JUN-1999; 99US-0139727.
XX PR 18-JUN-1999; 99US-0139728.
XX PR 18-JUN-1999; 99US-0139729.
XX PR 18-JUN-1999; 99US-0139730.
XX PR 18-JUN-1999; 99US-0139731.
XX PR 18-JUN-1999; 99US-0139732.
XX PR 18-JUN-1999; 99US-0139733.
XX PR 18-JUN-1999; 99US-0139734.
XX PR 18-JUN-1999; 99US-0139735.
XX PR 18-JUN-1999; 99US-0139736.
XX PR 18-JUN-1999; 99US-0139737.
XX PR 18-JUN-1999; 99US-0139738.
XX PR 18-JUN-1999; 99US-0139739.
XX PR 18-JUN-1999; 99US-0139740.
XX PR 18-JUN-1999; 99US-0139741.
XX PR 18-JUN-1999; 99US-0139742.
XX PR 18-JUN-1999; 99US-0139743.
XX PR 18-JUN-1999; 99US-0139744.
XX PR 18-JUN-1999; 99US-0139745.
XX PR 18-JUN-1999; 99US-0139746.
XX PR 18-JUN-1999; 99US-0139747.
XX PR 18-JUN-1999; 99US-0139748.
XX PR 18-JUN-1999; 99US-0139749.
XX PR 18-JUN-1999; 99US-0139750.
XX PR 18-JUN-1999; 99US-0139751.
XX PR 18-JUN-1999; 99US-0139752.
XX PR 18-JUN-1999; 99US-0139753.
XX PR 18-JUN-1999; 99US-0139754.
XX PR 18-JUN-1999; 99US-0139755.
XX PR 18-JUN-1999; 99US-0139756.
XX PR 18-JUN-1999; 99US-0139757.
XX PR 18-JUN-1999; 99US-0139758.
XX PR 18-JUN-1999; 99US-0139759.
XX PR 18-JUN-1999; 99US-0139760.
XX PR 18-JUN-1999; 99US-0139761.
XX PR 18-JUN-1999; 99US-0139762.
XX PR 18-JUN-1999; 99US-0139763.
XX PR 18-JUN-1999; 99US-0139764.
XX PR 18-JUN-1999; 99US-0139765.
XX PR 18-JUN-1999; 99US-0139766.
XX PR 18-JUN-1999; 99US-0139767.
XX PR 18-JUN-1999; 99US-0139768.
XX PR 18-JUN-1999; 99US-0139769.
XX PR 18-JUN-1999; 99US-0139770.
XX PR 18-JUN-1999; 99US-0139771.
XX PR 18-JUN-1999; 99US-0139772.
XX PR 18-JUN-1999; 99US-0139773.
XX PR 18-JUN-1999; 99US-0139774.
XX PR 18-JUN-1999; 99US-0139775.
XX PR 18-JUN-1999; 99US-0139776.
XX PR 18-JUN-1999; 99US-0139777.
XX PR 18-JUN-1999; 99US-0139778.
XX PR 18-JUN-1999; 99US-0139779.
XX PR 18-JUN-1999; 99US-0139780.
XX PR 18-JUN-1999; 99US-0139781.
XX PR 18-JUN-1999; 99US-0139782.
XX PR 18-JUN-1999; 99US-0139783.
XX PR 18-JUN-1999; 99US-0139784.
XX PR 18-JUN-1999; 99US-0139785.
XX PR 18-JUN-1999; 99US-0139786.
XX PR 18-JUN-1999; 99US-0139787.
XX PR 18-JUN-1999; 99US-0139788.
XX PR 18-JUN-1999; 99US-0139789.
XX PR 18-JUN-1999; 99US-0139790.
XX PR 18-JUN-1999; 99US-0139791.
XX PR 18-JUN-1999; 99US-0139792.
XX PR 18-JUN-1999; 99US-0139793.
XX PR 18-JUN-1999; 99US-0139794.
XX PR 18-JUN-1999; 99US-0139795.
XX PR 18-JUN-1999; 99US-0139796.
XX PR 18-JUN-1999; 99US-0139797.
XX PR 18-JUN-1999; 99US-0139798.
XX PR 18-JUN-1999; 99US-0139799.
XX PR 18-JUN-1999; 99US-0139800.
XX PR 18-JUN-1999; 99US-0139801.
XX PR 18-JUN-1999; 99US-0139802.
XX PR 18-JUN-1999; 99US-0139803.
XX PR 18-JUN-1999; 99US-0139804.
XX PR

PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159337.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	22-OCT-1999;	99US-0161404.
PR	23-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161922.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Search completed: July 25, 2001, 05:23:05
Job time: 4678 sec

JUL 25 10:29:06 2001

us-09-142-095-3.rml

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:18:52 ; Search time 117.39 seconds
(without alignments)
31.563 Million cell updates/sec

Title: US-09-142-095-3

Perfect score: 20

Sequence: 1 gtcacgtgacacgacgacaaac 20

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	16.4	82.0	22	3	US-08-857-464-3
C 2	14.8	74.0	519	4	US-08-991-789A-270
C 3	14.8	74.0	1990	2	US-08-967-101-154
C 4	14.8	74.0	1990	2	US-08-592-541-154
C 5	14.8	74.0	1990	2	US-09-124-698-154
C 6	14.8	74.0	1990	4	US-09-127-880-154
C 7	14.8	74.0	1990	4	US-08-496-841C-154
C 8	14.8	74.0	2740	1	US-08-100-247-4
C 9	14.8	74.0	2740	1	US-08-483-146A-4
C 10	14.8	74.0	2740	1	US-08-232-513A-5
C 11	14.8	74.0	2740	1	US-08-484-594A-4
C 12	14.4	72.0	448	1	US-09-060-756-666
C 13	14.2	71.0	84	1	US-08-053-131-145
C 14	14.2	71.0	84	1	US-08-645-641-145
C 15	14.2	71.0	84	1	US-07-853-408B-145
C 16	14.2	71.0	84	1	US-08-096-762-145
C 17	14.2	71.0	84	2	US-08-308-865-145
C 18	14.2	71.0	84	5	PCT-US92-10983-145
C 19	14.2	71.0	2700	3	US-09-022-875-1
C 20	13.8	69.0	410	4	US-09-060-756-239
C 21	13.8	69.0	453	4	US-09-060-756-189
C 22	13.8	69.0	1096	4	US-09-056-153-1
C 23	13.8	69.0	1941	4	US-09-402-002-1
C 24	13.8	69.0	2336	4	US-08-796-101-17
C 25	13.8	69.0	3022	4	US-08-961-083-215
C 26	13.8	69.0	4977	1	US-08-030-096-7
C 27	13.8	69.0	5275	4	US-08-796-101-19

C 28	13.8	69.0	35100	2	US-08-770-379-18	Sequence 18, Appl
C 29	13.8	69.0	35100	4	US-08-757-669A-18	Sequence 18, Appl
C 30	13.6	68.0	30	1	US-08-480-784-40	Sequence 40, Appl
C 31	13.6	68.0	30	1	US-08-483-553-40	Sequence 40, Appl
C 32	13.6	68.0	30	1	US-08-487-002-40	Sequence 40, Appl
C 33	13.6	68.0	30	1	US-08-483-554B-40	Sequence 40, Appl
C 34	13.6	68.0	30	1	US-08-488-011B-40	Sequence 40, Appl
C 35	13.6	68.0	30	5	US-08-850-727-40	Sequence 40, Appl
C 36	13.6	68.0	30	5	PCT-US95-10202-40	Sequence 40, Appl
C 37	13.6	68.0	30	5	PCT-US95-10202-40	Sequence 40, Appl
C 38	13.6	68.0	30	5	PCT-US95-10220-40	Sequence 40, Appl
C 39	13.6	68.0	372	4	US-09-124-671-30	Sequence 30, Appl
C 40	13.6	68.0	631	1	US-08-480-784-15	Sequence 15, Appl
C 41	13.6	68.0	631	1	US-08-483-553-15	Sequence 15, Appl
C 42	13.6	68.0	631	1	US-08-487-002-15	Sequence 15, Appl
C 43	13.6	68.0	631	1	US-08-483-554B-15	Sequence 15, Appl
C 44	13.6	68.0	631	1	US-08-488-011B-15	Sequence 15, Appl
C 45	13.6	68.0	631	4	US-08-850-727-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-857-464-3/c
Sequence 3, Application US/08857464
Patent No. 606450
GENERAL INFORMATION:
APPLICANT: Ralston, Stuart H.
APPLICANT: Grant, Stuart F.A.
TITLE OF INVENTION: Diagnostic and Therapeutic Methods and
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/857,464
FILING DATE: 16-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9610281.9
FILING DATE: 16-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1581.0180000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: both
MOLECULE TYPE: CDNA
US-08-857-464-3

Apparatus

Query Match 82.0% Score 16.4; DB 3; Length 22;
Best Local Similarity 94.4% Pred. No. 4.2;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 tcaacgtgacacgacgacaaac 19

Db 19 TCACGTGACACTCATAA 2

RESULT 2

US-08-991-789A-270/C
Sequence 270, Application US/08991789A
Patent No. 6225054

GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
Smith, John M.
Reed, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 292

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed IP Law Group

STREET: 701 Fifth Avenue, Suite 6300

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/991,789A

FILING DATE: 11-Dec-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Potter, Jane E. R.

REGISTRATION NUMBER: 33,332

REFERENCE/DOCKET NUMBER: 210121.419C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 270:

SEQUENCE CHARACTERISTICS:

LENGTH: 519 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 270:

Query Match 74.0%; Score 14.8; DB 4; Length 519;

Best Local Similarity 88.9%; Pred. No. 41;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcacgtgacacagtcacaa 19

Db 74 TCACGTGACACTCATAA 57

RESULT 3

US-08-967-101-154
Sequence 154, Application US/08967101
Patent No. 5840540

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMENS, JOHANNA M

APPLICANT: FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT

STREET: High Street Tower - 125 High Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/967,101

FILING DATE: 10-NOV-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/592,541

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 154:

SEQUENCE CHARACTERISTICS:

LENGTH: 1990 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-967-101-154

Query Match 74.0%; Score 14.8; DB 2; Length 1990;

Best Local Similarity 84.2%; Pred. No. 49;

Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tcacgtgacacagtcacaa 20

Db 168 TCACGTGACACTCATAA 186

RESULT 4

US-08-592-541-154
Sequence 154, Application US/08592541
Patent No. 5986054

GENERAL INFORMATION:

APPLICANT: ST. GEORGE-HYSLOP, PETER H

APPLICANT: ROMENS, JOHANNA M

APPLICANT: FRASER, PAUL E

TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED

TO ALZHEIMER'S DISEASE

NUMBER OF SEQUENCES: 183

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA, HURWITZ & THIBEAULT

STREET: High Street Tower - 125 High Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/592,541

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 154:

SEQUENCE CHARACTERISTICS:

LENGTH: 1990 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
;
US-08-552-541-154

```

Query Match	74.08;	Score 14.8;	DB 2;	Length 190;
Best Local Similarity	84.28;	Pred. No. 49;		
Matches 16; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

```

QY      2  tcaagtgacacagtcacac  20
          ||| ||||| ||| |||||
Db     168 TCAAGTGACCCAGNCAAC  186

```

RESULT 5
 US-09-124-698-154
 Sequence 154, Application US/09124698
 Patent No. 6117978
 GENERAL INFORMATION:
 APPLICANT: ST. GEORGE-HYSLOP, PETER H
 APPLICANT: ROMKENS, JOHANNA M
 APPLICANT: FRASER, PAUL E
 TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
 TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
 NUMBER OF SEQUENCES: 183
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: TESTA, HUMINTZ & THIBEAULT
 STREET: High Street Tower - 125 High Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.
 Zip: 02110
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION NUMBER: US/09/124,698

Query Match	74.0%;	Score 14.8;	DB 4;	Length 1990;
Best Local Similarity	84.2%;	Pred. No. 49;		

```

QY      2  tcacgtgacacagtcacac 20
          ||| ||||| ||| |||||
Db      168 tCAAGTGACCCAGNCAAC 186

```

RESULT 6
US-09-127-480-154
Sequence 154, Application US/09127480

```

Patent No. 6194153
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANN M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/127,480
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ. ID NO: 154:
SEQUENCE CHARACTERISTICS:
LENGTH: 1990 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-127-480-154

Query Match 74.0%; Score 14.8; DB 4; Length 1990;
Best Local Similarity 84.2%; Pred. NO. 49;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
0; 2 taactgacacgtcaac 20
|||||
168 tcgaatgacccagcnaaac 186

```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/496,841C

FILING DATE: 28-Jun-1995

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Paul F. Fehlner, Ph.D.

REGISTRATION NUMBER: 35,115

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 527-7700

TELEFAX: (212) 753-6237

INFORMATION FOR SEQ ID NO: 154:

SEQUENCE CHARACTERISTICS:

LENGTH: 1990 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: cDNA

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 154:

US-08-496-841C-154

Query Match 74.0%; Score 14.8; DB 4; Length 1990;
Best Local Similarity 84.2%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tcaagtgcacagtcacaa 20

Db 168 tcaagtgcacagtcacaa 186

RESULT 8
US-08-100-247-4

Sequence 4, Application US/08100247

Patent No. 5571787

GENERAL INFORMATION:

APPLICANT: O'BRIEN, JOHN S.

TITLE OF INVENTION: PROSAPOSIN AS A NEUROTROPHIC FACTOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: KNOBBE, MARTENS, OLSON AND BEAR

STREET: 620 NEWPORT CENTER DRIVE SIXTEENTH FLOOR

CITY: NEWPORT BEACH

STATE: CA

COUNTRY: USA

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/100,247

FILING DATE: 19930730

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Israelsen, Ned A.

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: O'BRIEN.002A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-235-8550

TELEFAX: 619-235-0176

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 2740 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

IMMEDIATE SOURCE:

CLONE: PROSAPOSIN cDNA

US-08-100-247-4

Query Match 74.0%; Score 14.8; DB 1; Length 2740;
Best Local Similarity 88.9%; Pred. No. 51;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcaagtgcacagtcacaa 19

Db 1194 tcaagtgcacagtcacaa 1211

RESULT 9
US-08-483-146A-4

Sequence 4, Application US/08483146A

Patent No. 5636080

GENERAL INFORMATION:

APPLICANT: O'Brien, John S.

TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS

TITLE OF INVENTION: COMPRISING PROSAPOSIN AND NEUROTROPHIC PEPTIDES DERIVED

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobb, Martens, Olson and Bear

STREET: 620 Newport Center Blvd. 16th Floor

CITY: Newport Beach

STATE: CA

COUNTRY: USA

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483,146A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Israelsen, Ned A.

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: MYELOS.002DV1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-235-8550

TELEFAX: 619-235-0176

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 2740 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-483-146A-4

Query Match 74.0%; Score 14.8; DB 1; Length 2740;

Best Local Similarity 88.9%; Pred. No. 51;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcaagtgcacagtcacaa 19

Db 1194 tcaagtgcacagtcacaa 1211

RESULT 10

US-08-232-513A-5

Sequence 5, Application US/08232513A

Patent No. 5700909
GENERAL INFORMATION:
APPLICANT: O'Brien, John S.
TITLE OF INVENTION: Prosaposin and Cytokine-derived peptides
TITLE OF INVENTION: as Therapeutic Agents
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232.513A
FILING DATE: 21-APR-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/100,247
FILING DATE: 30-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UD 1643
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2740 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2740
OTHER INFORMATION: /label=Hum_Prosaposin
US-08-232-513A-5

Query Match 74.0%; Score 14.8; DB 1; Length 2740;
Best Local Similarity 88.9%; Pred. No. 51;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcacgtgacacagtcacaa 19
|||||
DB 1194 TCACGTGACTGACGCAAA 1211

RESULT 11
US-08-484-594A-4
Sequence 4, Application US/08484594A
Patent No. 5714459
GENERAL INFORMATION:
APPLICANT: O'Brien, John S.
APPLICANT: Kishimoto, Yasuo
TITLE OF INVENTION: USE OF PROSAPOSIN AND NEUROTROPHIC PEPTIDES
TITLE OF INVENTION: DERIVED THEREFROM
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodder, Martens, Olson and Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fastseq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,594A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/100,247
FILING DATE: 30-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: MTELOS.0021V2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-235-8550
TELEFAX: 619-235-0176
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2740 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-484-594A-4

Query Match 74.0%; Score 14.8; DB 1; Length 2740;
Best Local Similarity 88.9%; Pred. No. 51;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcacgtgacacagtcacaa 19
|||||
DB 1194 TCACGTGACTGACGCAAA 1211

RESULT 12
US-09-060-756-666/C
Sequence 666, Application US/09060756
Patent No. 6183857
GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchliesser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 666
LENGTH: 448
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: unsure
LOCATION: (various positions within the sequence)
OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-666

Query Match 72.0%; Score 14.4; DB 4; Length 448;
Best Local Similarity 88.2%; Pred. No. 65;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 acgtgacacagtcacaa 20
|||||
DB 298 ACGGGACACAGTCGCAAC 282

RESULT 13

US-08-053-131-145/c

; Sequence 145, Application US/08053131

; Patent No. 5661016

; GENERAL INFORMATION:

; APPLICANT: Lomborg, Nils

; ATTORNEY/AGENT INFORMATION:

; TITLE OF INVENTION: Transgenic No. 5661016-Human Animals for

; TITLE OF INVENTION: Producing Heterologous Antibodies

; NUMBER OF SEQUENCES: 197

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend Kourile and Crew

; STREET: One Market Plaza, Stewart Tower, Suite 200

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94105

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/053,131

; FILING DATE: 26-APR-1993

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/990,860

; FILING DATE: 16-DEC-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/810,279

; FILING DATE: 17-DEC-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/853,408

; FILING DATE: 18-MAR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Smith, William M.

; REGISTRATION NUMBER: 30,223

; REFERENCE/DOCKET NUMBER: 14643-9-3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-326-2400

; TELEFAX: 415-326-2422

; INFORMATION FOR SEQ ID NO: 145:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 84 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-053-131-145

Query Match 71.0%; Score 14.2; DB 1; Length 84;

Best Local Similarity 84.2%; Pred. No. 66;

Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gtcacgtgacacagtcacaa 19

||||| | |||||

DB 36 gtcacgttaccagtcacaa 18

RESULT 14

US-08-645-641-145/c

; Sequence 145, Application US/08645641

; Patent No. 5719032

; GENERAL INFORMATION:

; APPLICANT: Lomborg, Nils

; ATTORNEY/AGENT INFORMATION:

; TITLE OF INVENTION: Transgenic No. 5719032-Human Animals for

; TITLE OF INVENTION: Producing Heterologous Antibodies

; NUMBER OF SEQUENCES: 150

CORRESPONDENCE ADDRESS:

ADDRESSEE: William M. Smith

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/645,641

; FILING DATE: 20-MAY-1996

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/904,068

; FILING DATE: 23-JUN-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Smith, William M.

; REGISTRATION NUMBER: 30,223

; REFERENCE/DOCKET NUMBER: 14643-000913

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-326-2400

; TELEFAX: 415-326-2422

; INFORMATION FOR SEQ ID NO: 145:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 84 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-645-641-145

Query Match 71.0%; Score 14.2; DB 1; Length 84;

Best Local Similarity 84.2%; Pred. No. 66;

Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gtcacgtgacacagtcacaa 19

||||| | |||||

DB 36 gtcacgttaccagtcacaa 18

RESULT 15

US-07-853-408B-145/c

; Sequence 145, Application US/07853408B

; Patent No. 5789650

; GENERAL INFORMATION:

; APPLICANT: Lomborg, Nils

; ATTORNEY/AGENT INFORMATION:

; TITLE OF INVENTION: Transgenic No. 5789650-Human Animals for

; TITLE OF INVENTION: Producing Heterologous Antibodies

; NUMBER OF SEQUENCES: 150

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: William M. Smith

; STREET: One Market Plaza, Stewart Tower, Suite 2000

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94105

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/853,408B

; FILING DATE: 19920318

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Smith, William M.

REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 14643-9
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ. ID NO.: 145:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-853-408-145

Query Match 71.0%; Score 14.2; DB 1; Length 84;
Best Local Similarity 84.2%; Pred. No. 66;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 gtacggtgacacagtcataa 19
||||| |
Db 36 gtacggttaaccagtcataa 18

Search completed: July 25, 2001, 05:18:53
Job time: 9216 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 04:54:51 ; Search time 2762.24 Seconds
(without alignments)
68.443 Million cell updates/sec

Title: US-09-142-095-3

Perfect score: 20

Sequence: 1 gtacagtgacacagctaac 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1022815 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*

190: gb_est110:*
 191: gb_est111:*
 192: gb_est112:*
 193: em_gss_fun:*
 194: em_gss_hum1:*
 195: em_gss_hum2:*
 196: em_gss_hum3:*
 197: em_gss_hum4:*
 198: em_gss_hum5:*
 199: em_gss_hum6:*
 200: em_gss_hum7:*
 201: em_gss_hum8:*
 202: em_gss_hum9:*
 203: em_gss_inv1:*
 204: em_gss_inv2:*
 205: em_gss_inv3:*
 206: em_gss_other:*
 207: em_gss_pln1:*
 208: em_gss_pln2:*
 209: em_gss_pro:*
 210: em_gss_r0d1:*
 211: em_gss_r0d2:*
 212: em_gss_r0d3:*
 213: em_gss_r0d4:*
 214: em_gss_r0d5:*
 215: em_gss_vrt1:*
 216: em_gss_vrt2:*
 217: em_gss_vrt3:*
 218: gb_gss1:*
 219: gb_gss2:*
 220: gb_gss3:*
 221: gb_gss4:*
 222: gb_gss5:*
 223: gb_gss6:*
 224: gb_gss7:*
 225: gb_gss8:*
 226: gb_gss9:*
 227: gb_gss10:*
 228: gb_gss11:*
 229: gb_gss12:*
 230: gb_gss13:*
 231: gb_gss14:*
 232: gb_gss15:*
 233: gb_gss16:*
 234: gb_gss17:*
 235: gb_gss18:*
 236: gb_gss19:*
 237: gb_gss20:*
 238: gb_gss21:*
 239: gb_gss22:*
 240: gb_gss23:*
 241: gb_gss24:*
 242: gb_gss25:*
 243: gb_gss26:*
 244: gb_gss27:*
 245: gb_gss28:*
 246: gb_gss29:*
 247: gb_gss30:*
 248: gb_gss31:*
 249: gb_gss32:*
 250: gb_gss33:*
 251: gb_gss34:*
 252: em_gss_inv4:*
 253: em_gss_r0d6:*
 254: em_gss_r0d7:*
 255: em_gss_r0d8:*
 256: gb_gss35:*
 257: gb_gss36:*
 258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18.4	92.0	737	166	BE306792	BE306792 601101950
2	18.4	87.0	238	130	BB310653	BB310653 BB310653
3	17.4	87.0	295	16	A1104756	A1104756 EST14045
4	17.4	87.0	438	237	A2015172	A2015172 RPCI-23-2
5	17.4	87.0	482	233	A0788101	A0788101 HS.3158.A
6	17.4	87.0	554	173	A1180191	A1180191 EST23332
7	16.8	84.0	258	222	FR0039113	AL126611 Fugu tubr
8	16.8	84.0	400	222	FR0039183	AL126681 Fugu tubr
9	16.8	84.0	513	222	FR0039084	AL126682 Fugu tubr
10	16.8	84.0	539	166	BE368176	BE368176 601222089
11	16.8	84.0	548	146	BF500440	BF500440 AT15369.5
12	16.8	84.0	562	190	W73212	W73212 z453610.f1
13	16.8	84.0	804	146	BF232638	BF232638 602023221
14	16.8	84.0	835	220	CNS0260P	AL195730 Tetradon
15	16.8	84.0	835	221	CNS03EDV	AL240368 Tetradon
16	16.8	84.0	968	220	CNS02RA3	AL210324 Tetradon
17	16.8	84.0	1046	172	BF982699	BF982699 602305085
18	16.8	84.0	1114	154	BG493528	BG493528 602542446
19	16.4	82.0	557	231	AC657757	AC657757 Sheared D
20	16.4	82.0	619	222	FR007351	Z9161.F.rudripes
21	16.4	82.0	837	221	CNS03PON	AL254984 Tetradon
22	16.4	82.0	873	150	BF578047	BF578047 602091895
23	16.4	82.0	892	220	CNS02OLQ	AL15751 Tetradon
24	16.4	82.0	988	219	CNS01PDJ	AL154679 Anopheles
25	16.4	82.0	971	221	CNS04LBN	AL286060 Tetradon
26	16.4	82.0	981	221	CNS04LBN	AL274375 Tetradon
27	15.8	79.0	181	250	AZ841578	AZ841578 2M0139F09
28	15.8	79.0	199	127	BB160957	BB160957 BB160957
29	15.8	79.0	353	152	BG318736	BG318736 NXYE.016
30	15.8	79.0	264	29	AV329412	AV329412 AV329412
31	15.8	79.0	279	109	AV062353	AV062353 AV062353
32	15.8	79.0	284	25	AV158238	AV158238 AV158238
33	15.8	79.0	287	16	AV111900	AV111900 UT-R-Al-e
34	15.8	79.0	303	14	AA957109	AA957109 UT-R-EI-E
35	15.8	79.0	311	15	AI040776	AI040776 OX15h08.S
36	15.8	79.0	327	17	AI228082	AI228082 EST224777
37	15.8	79.0	345	22	AI598772	AI598772 EST251475
38	15.8	79.0	370	224	AQ130228	AQ130228 HS.3007.A
39	15.8	79.0	375	112	AM122511	AM122511 EST292763
40	15.8	79.0	381	114	AM295307	AM295307 UT-H-812-
41	15.8	79.0	387	122	AM919078	AM919078 EST350382
42	15.8	79.0	400	188	R92042	R92042 JP96604.r1
43	15.8	79.0	404	187	R17651	R17651 YG15608.r1
44	15.8	79.0	409	14	AA944486	AA944486 EST199985
45	15.8	79.0	420	12	AA800910	AA800910 EST190407

ALIGNMENTS

RESULT 1
 BE306792/c
 LOCUS
 DEFINITION
 BE306792 737 bp mRNA
 BE306792 601101950P1 NC1_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3494340 5',
 mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 (bases 1 to 737)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 Mammalian Gene Collection (MGC)
 UNPUBLISHED (1999)
 CONTACT: Robert Strausberg, Ph.D.

FEATURES

SOURCE

Email: cgapds-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LLM8543 row: c column: 13
 High quality sequence stop: 359.
 Location/Qualifiers
 1..737
 /organism="Mus musculus"
 /strain="C57BL/6J (f1er1)"
 /db_xref="taxon:10090"
 /clone="IMAGE:3494340"
 /clone_id="NC1_CGAP_Lu29"
 /tissue_type="spontaneous tumor, metastatic to mammary."
 /stem_cell_origin=""
 /lab_host="DH10B"
 /note="Organ: Lung; Vector: pCMV-SPORT6; Site: 1: SalI;
 Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT

169 a 190 c 177 g 201 t

Query Match

Best Local Similarity 92.0%; Score 18.4; DB 166; Length 737;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtcacgtgacacagtcacac 20
 DB 305 GTCACGTGACACGTACAC 286

RESULT 2
 BE310653/c
 LOCUS
 DEFINITION
 BE310653 RIKEN full-length enriched adult male corpora
 quadrigemina Mus musculus cDNA clone B230316K05 3', mRNA sequence.
 BE310653
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 (bases 1 to 238)
 Kono H., Aizawa K., Akahira S., Akiyama J., Arakawa T., Carninci
 P., Endo T., Fukuda S., Fukunishi Y., Hara A., Hayatsu N.,
 Hirozane T., Hori F., Ishii Y., Ishikawa J., Ishikawa T., Itoh M.,
 Izawa M., Kadota K., Kagawa I., Kai C., Kawada C., Kusabe M.,
 Kiyosawa H., Kojima Y., Kondo S., Koya S., Kurihara C., Kusabe M.,
 Matsuyama T., Miki R., Mizuno Y., Nakamura M., Oda H., Okazaki Y.,
 Ono T., Owa C., Saito H., Sakai C., Saito K., Shibata K., Shibata
 Y., Shigemoto Y., Shinagawa A., Shiraki T., Sogabe Y., Sugahara Y.,
 Suzuki H., Suzuki H., Tagawa A., Takahashi F., Tomimaga N., Toyota
 T., Tsunoda Y., Watanabe S., Yamada S., Yamada T., Yamashita I.,
 Yano R., Yasunishi A., Yokota T., Yoshida K., Yoshiki A., Yoshino
 M., Muramatsu M. and Hayashizaki Y.
 RIKEN House ESTs (Kono H., et al.)
 UNPUBLISHED (2000)
 CONTACT: Yoshihide Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 Genome Science Laboratory
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-research@riken.go.jp

URL: <http://genome.rtc.riken.go.jp/>
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagata, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Thermostabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)
 Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Iwaw, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)
 Carninci, P., and Hayashizaki, Y.
 High efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp/>) for further details.

FEATURES

source

Location/Qualifiers

1. 238
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="B230316K05"
 /clone_id="RIKEN full-length enriched, adult male corpora quadrigemina"
 /sex="male"
 /tissue_type="corpora quadrigemina"
 /dev_stage="adult"
 /lab_host="DH10B"

/note="Site 1: Salt; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot - 20.0 and subtraction to Rot - 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGATTCGATGATTAATTAATCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX I."

BASE COUNT

71 a 43 c 51 g 73 t

ORIGIN

Query Match

Best Local Similarity 87.0%; Score 17.4; DB 130; Length 238;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtcacgtgacacagtcacaa 19
 |||||||
 Db 72 gtcacgtgacacagtcacaa 54

RESULT 3

A1104756

295 bp mRNA

EST

20-JAN-1999

LOCUS

DEFINITION

EST211045 Normalized rat heart, Bento Soares Rattus sp. cDNA clone

ACCESSION

A1104756

VERSION

A1104756.1 GI:3709041

KEYWORDS

EST

SOURCE

Rattus sp.

ORGANISM

Rattus sp.
 Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE

1 (bases 1 to 295)

AUTHORS

Lee, N.H., Glodex, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Keriavage, A.R. and Adams, M.D.

TITLE

JOURNAL

unpublished (1998)

COMMENT

Gene Index
 Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
 Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: ML3-21.

FEATURES

source

Location/Qualifiers

1. 295
 /organism="Rattus sp."
 /db_xref="ATCC (lnhost):2025643"
 /db_xref="taxon:10118"
 /clone="RHEC160"
 /clone_id="Normalized rat heart, Bento Soares"
 /note="Organ: heart; Vector: pT73Pac; Site 1: EcoRI;
 Site 2: NotI"

BASE COUNT

87 a 49 c 74 g 85 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 16; Length 295;
 Best Local Similarity 94.7%; Pred. No. 92;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tcaacgtgacacagtcacaa 20
 |||||||
 Db 233 TCACGTGACACACATCAAC 251

RESULT 4

A2015172

438 bp DNA

GSS 25-FEB-2000

LOCUS

RPCI-23-289M15.TJ RPCI-23 Mus musculus genomic clone RPCI-23-289M15

DEFINITION

DNA sequence.

ACCESSION

A2015172

VERSION

A2015172.1 GI:7090556

KEYWORDS

GSS

SOURCE

house mouse.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 438)

AUTHORS

Zhao, S., Nierman, M., Feldblum, T., Malek, J., Shatsman, S., Aklurel,
 B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.

TITLE

Mouse BAC End Sequences from Library RPCI-23

JOURNAL

unpublished (1999)

COMMENT

Other GSSs: RPCI-23-289M15.TV

Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@tigr.org, med.bufileo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.bufileo.edu/orderingframe.htm>)
 or from Resea ch Genetics (<http://resch.genetics.com>). BAC end page:
http://www.tigr.org/tidb/bac/ends/mouse/bac_end_intro.html
 Plate: 289 row: M Column: 15
 Seq primer: SP6
 Class: BAC ends.

FEATURES

source

Location/Qualifiers

1. 438
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-289M15"

Umranta, Y., Williams, G. and Brenner, S.
Direct Submission
Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
Centre, Hinxton, Cambridge, CB10 1SB, UK Email:

COMMENT

biohelphgmp.mrc.ac.uk
Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR:

One pass dye-terminator sequencing of cosmid cloned genomic sequence.

FEATURES

SOURCE

Location/Qualifiers
1..258

/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_1lb="cosmid 090H03"
/clone="090H03d37"

BASE COUNT 67 a 66 c 58 g 64 t 3 others
ORIGIN

Query Match

84.0%; Score 16.8; DB 222; Length 258;
Best Local Similarity 90.0%; Pred. No. 1.9e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gtcaagtgcacagtcacaac 20
||||| ||||||| |||||

Db 200 gtcacatgacacagtcacac 181

RESULT 8

FR0039183/c

LOCUS FR0039183 400 bp DNA 22-OCT-1999

DEFINITION Fugu rubripes GSS sequence, clone 090H03d3, genomic survey sequence.

ACCESSION AL126681.1 GI:6108296

VERSION GSS; genome survey sequence.

KEYWORDS Takifugu rubripes.

SOURCE Takifugu rubripes.

ORGANISM Takifugu rubripes.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.

REFERENCE 1 (bases 1 to 400)

AUTHORS Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,

Umranta,Y., Williams,G. and Brenner,S.

TITLE Direct Submission

JOURNAL Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource

Centre, Hinxton, Cambridge, CB10 1SB. UK Email:

biohelphgmp.mrc.ac.uk

Vector: pBluescript II KS

V_type: phagemid

PRIMER: KS

DESCR: One pass dye-terminator sequencing of cosmid cloned genomic

sequence.

FEATURES

SOURCE

Location/Qualifiers
1..400

/organism="Takifugu rubripes"

/db_xref="taxon:31033"

/clone_1lb="cosmid 090H03"

/clone="090H03d37"

BASE COUNT 102 a 102 c 88 g 103 t 5 others
ORIGIN

Query Match

84.0%; Score 16.8; DB 222; Length 400;
Best Local Similarity 90.0%; Pred. No. 2e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gtcaagtgcacagtcacaac 20
||||| ||||||| |||||

Db 296 gtcacatgacacagtcacac 277

RESULT 9

FR0039084/c

LOCUS FR0039084 513 bp DNA

DEFINITION Fugu rubripes GSS sequence, clone 090H03ca5, genomic survey

sequence.

ACCESSION AL126582.1 GI:6108197

VERSION GSS; genome survey sequence.

KEYWORDS Takifugu rubripes.

SOURCE Takifugu rubripes.

ORGANISM Takifugu rubripes.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.

REFERENCE 1 (bases 1 to 513)

AUTHORS Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,

Umranta,Y., Williams,G. and Brenner,S.

TITLE Direct Submission

JOURNAL Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource

Centre, Hinxton, Cambridge, CB10 1SB. UK Email:

biohelphgmp.mrc.ac.uk

Vector: pBluescript II KS

V_type: phagemid

PRIMER: KS

DESCR: One pass dye-terminator sequencing of cosmid cloned genomic

sequence.

FEATURES

SOURCE

Location/Qualifiers
1..513

/organism="Takifugu rubripes"

/db_xref="taxon:31033"

/clone_1lb="cosmid 090H03"

/clone="090H03ca5"

BASE COUNT 132 a 130 c 112 g 134 t 5 others
ORIGIN

Query Match

84.0%; Score 16.8; DB 222; Length 513;
Best Local Similarity 90.0%; Pred. No. 2e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gtcaagtgcacagtcacaac 20
||||| ||||||| |||||

Db 30 gtcacatgacacagtcacac 11

RESULT 10

BE368176

LOCUS BE368176 539 bp mRNA 21-JUL-2000

DEFINITION 601222089F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3590711 5',

mRNA sequence.

ACCESSION BE368176

VERSION BE368176.1 GI:9313539

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 539)

AUTHORS NIH-MGC http://mgc.ncl.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs@remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

plate: LAM8759 row: b column: 24

High quality sequence stop: 536.

Location/Qualifiers

FEATURES

```

source
1..539
/organism="Mus musculus"
/strain="C57BL/6J (f1)"
/db_xref="taxon:10090"
/clone="IMAGE:3590711"
/clone_1ib="NCI_CGAP_Lu29"
/issue_type="spontaneous tumor, metastatic to mammary.
stem cell origin."
/lab_host="DR108"
/site="Organ: lung; Vector: pCMV-SPORT6; Site: 1; Sall;
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
201 a 92 c 143 g 103 t

BASE COUNT
ORIGIN

Query Match 84.0%; Score 16.8; DB 166; Length 539;
Best Local Similarity 90.0%; Pred. No. 2.1e-02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtcacgtgacgcagtcacac 20
|||||
Db 207 gtcacgtgacgcagtcacac 226

RESULT 11
BF500440/c 548 bp mRNA EST 16-APR-2001
LOCUS AT15369 5prime AT Drosophila melanogaster adult testes POTB7
DEFINITION Drosophila melanogaster cDNA AT15369 5 similar to CG9778:
Phan0009778 lastc_updated:000321. mRNA sequence.
BF500440
BF500440.1 GI:11583741
EST.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 548)
Stapleton, M., Brokstein, P., Hong, L., Agdayani, A., Baxter, E., Berman,
B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan,
D., Faise, E., George, R., Gonzalez, M., Garin, H., Harris, N., Li, P.,
Liao, G., Miranda, A., Mistre, S., Mungall, C. J., Nunoo, J., Pacleb, J.,
Park, S., Paragas, A., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E.,
Celniker, S. and Rubin, G. M.
Berkeley Drosophila Gene Collection Project
Unpublished (2000)
Contact: Stapleton, M.
BDGP

TITLE
JOURNAL
COMMENT

FEATURES
SOURCE
1..548
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1ib="AT15369"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT-10-AT-120: DHS-alpha. Plates
AT-121-AT-319: DHS-alpha Tona"
/notes="Organ: ADULT testes; Vector: pOTB7; Site: 1; Scori;
Site: 2; XhoI; The mRNA for the testis library was made
from testes and seminal vesicles and dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly

```

[illegible]

Euharptia; Merazoa; Choriata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 835)

AUTHORS Genoscope.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source

1.835
location/Qualifiers
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="020G19"
/clone_lib="G"
/note="Genoscope sequence ID : C08G020AD10SP1-end :
pUC-ori"

BASE COUNT 185 a 245 c 233 g 168 t 4 others

ORIGIN

Query Match 84.0%; Score 16.8; DB 221; Length 835;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 gtcaagtgacacagtcacac 20
|||||
Db 261 gtcaagtgacacagtcacac 280

Search completed: July 25, 2001, 04:54:54
Job time: 10427 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:17:09 ; Search time 1290.33 Seconds
(without alignments)
227.761 Million cell updates/sec

Title: US-09-142-095-4

Perfect score: 19
Sequence: 1 ttgtctctccagaggtt 19

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenDbml:**

1: gb_da1:*
2: gb_da2:*
3: gb_da3:*
4: gb_da4:*
5: gb_da5:*
6: gb_da6:*
7: gb_da7:*
8: gb_da8:*
9: gb_da9:*
10: gb_da10:*
11: gb_da11:*
12: gb_da12:*
13: gb_da13:*
14: gb_da14:*
15: gb_da15:*
16: gb_da16:*
17: gb_da17:*
18: gb_da18:*
19: gb_da19:*
20: gb_da20:*
21: gb_da21:*
22: gb_da22:*
23: gb_da23:*
24: gb_da24:*
25: gb_da25:*
26: gb_da26:*
27: gb_da27:*
28: gb_da28:*
29: gb_da29:*
30: gb_da30:*
31: gb_da31:*
32: gb_da32:*
33: gb_da33:*
34: gb_da34:*
35: gb_da35:*
36: gb_da36:*
37: gb_da37:*
38: gb_da38:*
39: gb_da39:*
40: gb_da40:*
41: gb_da41:*
42: gb_da42:*
43: gb_da43:*

44: em_ov:*
45: em_ph:*
46: em_pi:*
47: em_pi:*
48: em_pi:*
49: em_pi:*
50: em_pi:*
51: em_pi:*
52: em_pi:*
53: gb_da1:*
54: gb_da2:*
55: gb_da3:*
56: gb_da4:*
57: gb_da5:*
58: gb_da6:*
59: gb_da7:*
60: gb_da8:*
61: gb_da9:*
62: gb_da10:*
63: gb_da11:*
64: gb_da12:*
65: gb_da13:*
66: gb_da14:*
67: gb_da15:*
68: gb_da16:*
69: gb_da17:*
70: gb_da18:*
71: gb_da19:*
72: gb_da20:*
73: gb_da21:*
74: gb_da22:*
75: gb_da23:*
76: gb_da24:*
77: gb_da25:*
78: gb_da26:*
79: gb_da27:*
80: gb_da28:*
81: gb_da29:*
82: gb_da30:*
83: gb_da31:*
84: gb_da32:*
85: gb_da33:*
86: gb_da34:*
87: gb_da35:*
88: gb_da36:*
89: gb_da37:*
90: gb_da38:*
91: gb_da39:*
92: gb_da40:*
93: gb_da41:*
94: gb_da42:*
95: gb_da43:*
96: gb_da44:*
97: gb_da45:*
98: gb_da46:*

pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match Length	ID	Description
1	19	100.0	9	A65503 Sequence 4
2	19	100.0	89	AF135465 Gorilla g
3	19	100.0	200	AF135466 Pongo pyg
4	19	100.0	207	AF135470 Cebus ape
5	19	100.0	208	AF135464 Gorilla g
6	19	100.0	208	AF135471 Salimiri b
7	19	100.0	234	AF135462 Pan panis
8	19	100.0	89	AF135463 Pan trogl

```

c 9 19 100.0 292 89 AF357220 Homo sapi
c 10 19 100.0 531 89 AF352795 Homo sapi
c 11 19 100.0 541 89 AF180372 Homo sapi
c 12 19 100.0 620 9 AF35304
c 13 19 100.0 918 88 AF10194 Homo sapi
c 14 19 100.0 1190 97 HUMGT1A
c 15 19 100.0 3341 91 DB7674
c 16 19 100.0 68770 86 AC006985
c 17 19 100.0 176619 80 AC026497
c 18 19 100.0 198872 89 AF297093
c 19 17.4 91.6 208 89 AF133467
c 20 17.4 91.6 208 89 AF133468
c 21 17.4 91.6 113347 76 AC083820
c 22 17.4 91.6 158091 87 AC018904
c 23 17.4 91.6 155730 72 AC061988
c 24 17 88.5 159762 62 AC012040
c 25 17 89.5 150184 88 AC026464
c 26 17 89.5 196518 70 AC026474
c 27 17 89.5 208378 76 AC079988
c 28 16.4 86.3 910 1 AF134384
c 29 16.4 86.3 920 2 AF293014
c 30 16.4 86.3 920 2 AF293015
c 31 16.4 86.3 1338 3 UEAJ9508
c 32 16.4 86.3 1340 3 UEAJ9509
c 33 16.4 86.3 1350 3 MSU89773
c 34 16.4 86.3 1352 2 AF229776
c 35 16.4 86.3 1362 3 MVU20150
c 36 16.4 86.3 1420 3 MVU20154
c 37 16.4 86.3 1421 3 MMU20151
c 38 16.4 86.3 1422 3 MTHRR165
c 39 16.4 86.3 1423 3 MTHRR166
c 40 16.4 86.3 1425 3 MESRR165B
c 41 16.4 86.3 1426 3 MESRR165
c 42 16.4 86.3 1426 3 MESRR165
c 43 16.4 86.3 1426 3 MESRR165
c 44 16.4 86.3 1427 3 MESRR165A
c 45 16.4 86.3 1428 3 MESRR165G

```

ALIGNMENTS

```

RESULT 1
LOCUS A65503 19 bp DNA
DEFINITION Sequence 4 from Patent WO9732042.
ACCESSION A65503
VERSION A65503.1 GI:4531238
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 19)
AUTHORS Burchell, B.
TITLE DRUG TRIAL ASSAY SYSTEM
JOURNAL Patent: WO 9732042-A 4 04-SEP-1997;
UNIV DUNDIE (GB)
COMMENT Other publication AU 2224197 19970916.
FEATURES
source
1. 19
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 2 a 5 c 5 g 7 t
ORIGIN

```

```

Query Match 100.0%; Score 19; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtctctgcagaggtt 19
db 1 ttgtctctgcagaggtt 19

```

```

RESULT 2
LOCUS AF133465/c
DEFINITION AF133465 177 bp DNA
ACCESSION AF133465
VERSION AF133465
KEYWORDS
SOURCE
ORGANISM
TITLE
REFERENCE 1 (bases 1 to 177)
AUTHORS Hall, D., Ylazeta, G., Destro-Bisoli, G., Petzi-Erler, M.L. and Di Rienzo, A.
VARIABILITY at the uridine diphosphate glucuronosyltransferase 1A1 promoter in human populations and primates
Pharmacogenetics (1999) In press
JOURNAL
REFERENCE 2 (bases 1 to 177)
AUTHORS Ylazeta, G., Hall, D. and Di Rienzo, A.
Direct Submision
Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924 E. 57th Street, Chicago, IL 60637, USA
LOCATION/Qualifiers
1. 177
/organism="Gorilla gorilla"
/db_xref="taxon:9593"
/note="from the Brookfield Zoo"
<91. >177
/gene="UGT1A1"
/product="UDP-glucuronosyltransferase 1A1"
<91. >177
/gene="UGT1A1"
<91. >177
/gene="UGT1A1"
/codon_start=1
/product="UDP-glucuronosyltransferase 1A1"
/db_xref="taxon:9593"
/db_xref="GI:6456548"
/translation="MAVESQGGHPVYLCLLCVGPVSHAK"

```

```

BASE COUNT 37 a 39 c 57 g 44 t
ORIGIN
Query Match 100.0%; Score 19; DB 89; Length 177;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtctctgcagaggtt 19
db 84 ttgtctctgcagaggtt 66

```

```

RESULT 3
LOCUS AF133466/c
DEFINITION AF133466 200 bp DNA
ACCESSION AF133466
VERSION AF133466
KEYWORDS
SOURCE
ORGANISM
TITLE
REFERENCE 1 (bases 1 to 200)
AUTHORS Hall, D., Ylazeta, G., Destro-Bisoli, G., Petzi-Erler, M.L. and Di Rienzo, A.
VARIABILITY at the uridine diphosphate glucuronosyltransferase 1A1 promoter in human populations and primates
Pharmacogenetics (1999) In press
JOURNAL

```

BASE COUNT	43 a	46 c	62 g	56 t
ORIGIN				
Query Match		100.0%;	Score 19;	DB 89;
Best Local Similarity		100.0%;	Pred. No. 15;	Length 207;
Matches	19;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

```

0Y      1 ttgtctccgcagaggtt 15
          |||||
Db      114 TTGCTCCTGCCAGAGGT 96

```

RESULT	5
AF135464/c	
LOCUS	AF135464
DEFINITION	Gorilla gorilla UDP-glucuronosyltransferase 1A1 (UGT1A1) gene promoter region and partial cds.
ACCESSION	AF135464
VERSION	AF135464.1
KEYWORDS	GI:6456545
SOURCE	gorilla.
ORGANISM	Gorilla gorilla

REFERENCE	AUTHORS	TITLE
1 (pages 1 to 208)	Hall, D., Yazdeta, G., Desiro-Bisoli, G., Petzl-Erler, M. L. and Di Rienzo, A.	Variability at the uridine diphosphate glucuronosyltransferase 1A1 locus in Argentinean and estimates

FILE
JOURNAL
 Direct download
 Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
 E. 57th street, Chicago, IL 60637, USA
FEATURES
 Location/Qualifiers
 1.208

```

mRNA      <133..>208
           /gene="UCR1A1"
           /product="UDP-glucuronosyltransferase 1A1"

```

gene	133. :2408
CDS	/gene="UGT1A1" 133. :208

```

/codon_start=1
/product="UDP-glucuronosyltransferase 1A1"
/protein_id="AAF09175.1"

```

BASE COUNT	43 a	47 c	62 g	50 l
ORIGIN				

Query Match	100.0%;	Score 19;	DB 63;	Length 206;
Best Local Similarity	100.0%;	Pred. No. 15;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Db 126 TTTGCTCCCTGCCAGAGGTT 108

RESULT	6
AF135471/c	
LOCUS	AF135471 208 bp DNA PRI 21-NOV-1999
DEFINITION	Salmidri boliviensis UDP-glucuronosyltransferase 1A1 (UGT1A1) gene, promoter region and partial cdd.
ACCESSION	AF135471
VERSION	AF135471.1 GI:6456559

KEYWORDS

Bolivian squirrel monkey.

SOURCE

Salmiti boliviensis

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae;

REFERENCE

1 (bases 1 to 208)

AUTHORS

TITLE

Variability at the uridine diphosphate glucuronosyltransferase 1A1

JOURNAL

Pharmacogenetics (1999) In press

AUTHORS

Ybazaeta,G., Hall,D. and Di Rienzo,A.

TITLE

Direct Submission

JOURNAL

Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924

FEATURES

E. 57th Street, Chicago, IL 60637, USA

source

Location/Qualifiers

mrna

1. 208

gene

/organism="Salmiti boliviensis"

CDS

/db_xref="taxon:27679"

BASE COUNT

44 a 52 c 58 g 54 t

ORIGIN

Query Match 100.0%; Score 19; DB 89; Length 208;

Best Local Similarity 100.0%; Pred. No. 15; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtctctgcagaggtt 19

Db 135 TTGCTCTGCAGAGGTT 117

RESULT 7

AF135462/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

Db 141 TTTCCTCTGCAGAGTT 123

RESULT 9
AF357220/c 292 bp DNA PRI 27-MAR-2001
LOCUS Homo sapiens bilirubin UDP-glucuronosyltransferase 1 (UGT1) gene,
DEFINITION UGT1*1 allele, promoter and partial cds.
ACCESSION AF357220
VERSION AF357220.1 GI:13448828
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 292)
AUTHORS McKie,K., Addington,T., Nguyen,T.S., Glendenning,M., Kutlar,F. and
Kutlar,A.
TITLE Detection of TATA box 7A repeat region [6(TA)repeat] of human
bilirubin UDP-glucuronosyltransferase 1-1 gene (UGT1*1) in an
African American individual
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 292)
AUTHORS McKie,K., Addington,T., Nguyen,T.S., Glendenning,M., Kutlar,F. and
Kutlar,A.
TITLE Direct Submission
SUBMITTED (07-MAR-2001) Medicine/Hemoglobin DNA Laboratory, Medical
College of Georgia, 15th Street, Augusta, GA 30912, USA
LOCATION/Qualifiers
1..292

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37"
/cell_type="WBC"
/tissue_type="whole blood"
1..>292
/gene="UGT1"
/allele="UGT1*1"
1..64
/gene="UGT1"
28..39
/gene="UGT1"
/note="most common type of repeat; 6(TA)"
28..42
TATA_signal
28..39
/gene="UGT1"
repeat_region
/note="polymorphic region"
/rpt_type=tandem
/rpt_unit=ta
65..>292
/gene="UGT1"
/product="bilirubin UDP-glucuronosyltransferase 1"
81..>292
/codon_start=1
/product="bilirubin UDP-glucuronosyltransferase 1"
/protein_id="AAK27223.1"
/db_xref="GI:13448829"
/translation="MAVESOGGRPLVGLLCLVLPVSHAGKILLIPVDGSHMLSL
GAIQDLQGRHEIYVLAAPDASIVRPG"

BASE COUNT 60 a 69 c 95 g 68 t
ORIGIN

Query Match 100.0%; Score 19; DB 89; Length 292;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtctctgcagaggtt 19
|||||
Db 74 TTTCCTCTGCAGAGTT 56

RESULT 10
AF352795/c 531 bp DNA PRI 10-APR-2001
LOCUS Homo sapiens bilirubin UDP-glucuronosyltransferase 1-1 (UGT1A1) gene,
DEFINITION UGT1A1*1 allele, partial cds.
ACCESSION AF352795
VERSION AF352795.1 GI:13569708
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 531)
AUTHORS McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
TITLE 7(TA) repeat polymorphism of the TATA box of human bilirubin
UDP-glucuronosyltransferase 1-1(UGT1A1*1) gene in a patient with
sickle cell anemia + high bilirubinemia
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 531)
AUTHORS McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
TITLE Direct Submission
SUBMITTED (23-FEB-2001) Medicine/Hemoglobin DNA Laboratory, Medical
College of Georgia, 15th St., AC-1000, Augusta, GA 30912, USA
LOCATION/Qualifiers
1..531

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37"
/cell_type="white blood cell"
/tissue_type="blood"
1..>531
/gene="UGT1A1"
/note="UGT1"
/allele="UGT1A1*1"
1..313
/gene="UGT1A1"
279..295
/gene="UGT1A1"
279..295
/gene="UGT1A1"
/note="Repeat polymorphism compared to UGT1A1 sequence
presented in GenBank Accession Number AF180372; contains 7
ta repeats frequently detected in patients with high
bilirubinemia"
/rpt_type=tandem
/rpt_unit=ta
314..>531
/gene="UGT1A1"
/product="bilirubin UDP-glucuronosyltransferase 1-1"
334..>531
/gene="UGT1A1"
/codon_start=1
/product="bilirubin UDP-glucuronosyltransferase 1-1"
/protein_id="AAK31204.1"
/db_xref="GI:13569709"
/translation="MAVESOGGRPLVGLLCLVLPVSHAGKILLIPVDGSHMLSL
GAIQDLQGRHEIYVLAAPDASIVRPG"

BASE COUNT 120 a 121 c 137 g 153 t
ORIGIN

Query Match 100.0%; Score 19; DB 89; Length 531;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtctctgcagaggtt 19
|||||
Db 327 TTTCCTCTGCAGAGTT 309

RESULT 11

AF180372/c 541 bp DNA PRI 05-OCT-1999
 LOCUS Homo sapiens bilirubin UDP-glucuronosyltransferase 1-1 (UGT1) gene,
 DEFINITION UGT1*1 allele, partial cds.
 ACCESSION AF180372
 VERSION AF180372.1 GI:6010649
 KEYWORDS

SOURCE

ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 541)
 Kullar, F., Stomak, E., Leitner, C., Nechtman, J. and Kullar, A.
 Detection of the TATA box polymorphism of the human bilirubin
 UDP-glucuronosyltransferase 1-1 gene (UGT1*1) in a patient with
 sickle cell anemia

JOURNAL

unpublished

REFERENCE

2 (bases 1 to 541)
 Kullar, F., Stomak, E., Leitner, C., Nechtman, J. and Kullar, A.
 Submitted (24-AUG-1999) Medicine, Hematology/Oncology-Sickle Cell
 Center, Medical College of Georgia, 15th Street, AC-1000, Augusta,
 GA 30912, USA

FEATURES

source
 1..541
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2q37"
 /cell_type="white blood cells"
 /tissue_type="whole blood"
 1..>541
 /gene="UGT1"
 /note="GNT1; UGT1A1"
 /allele="UGT1*1"
 1..313
 /gene="UGT1"
 326
 /gene="UGT1"
 /replace="t"
 279..288
 /note="Polymorphic region"
 /rpl_type="tandem"
 repeat_region
 279..291
 /gene="UGT1"
 /note="Ritter, J.K., et al., 1992, J. Biol. Chem.,
 267:3257-3261"
 314..>541
 /gene="UGT1"
 /product="bilirubin UDP-glucuronosyltransferase 1-1"
 330..>541
 /gene="UGT1"
 /note="UDP glucosyltransferase 1"
 /codon_start=-1
 /product="bilirubin UDP-glucuronosyltransferase 1-1"
 /protein_id="AF01205.1"
 /db_xref="GI:6010650"
 /translation="MAVESQGRPLVGLLTVLPVYSHACKILLIPVDSHMLSL
 GAIQLOQRHEIVLAPASLYIDG"

CDS

BASE COUNT 123 a 124 c 141 g 153 t
 ORIGIN

Query Match 100.0%; Score 19; DB 89; Length 541;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtctctgcagaggtt 19
 |||||
 Db 323 TTTCCTCTGCCAGAGGTT 305

RESULT 12

AF180372/c 620 bp DNA PAT 29-MAR-1999
 LOCUS A65504
 DEFINITION Sequence 5 from Patent WO9732042.
 ACCESSION A65504
 VERSION A65504.1 GI:4531239
 KEYWORDS

SOURCE

ORGANISM unidentified.
 unclassified.

REFERENCE 1 (bases 1 to 620)
 Burchell, B.
 TITLE DRUG TRIAL ASSAY SYSTEM
 JOURNAL PATENT: WO 9732042-A 5 04-SEP-1997;
 UNIT DUNDEE (GB)
 COMMENT Other publication AU 2224197 19970916.
 FEATURES
 source
 1..620
 /organism="unidentified"
 /db_xref="taxon:32644"

BASE COUNT 157 a 127 c 151 g 185 t
 ORIGIN

Query Match 100.0%; Score 19; DB 9; Length 620;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtctctgcagaggtt 19
 |||||
 Db 605 TTTCCTCTGCCAGAGGTT 587

RESULT 13

AF110194/c 918 bp DNA PRI 02-JAN-2001
 LOCUS AF110194
 DEFINITION Homo sapiens chromosome 2 UDP-glucuronosyltransferase (UGT1A1)
 ACCESSION AF110194
 VERSION AF110194.1 GI:12002134
 KEYWORDS

SOURCE

ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 918)
 Gullerme, C.
 TITLE Direct Submission
 JOURNAL Submitted (01-DEC-1998) Center for Cancer Research, MIT, 77
 Massachusetts Avenue, E17-540, Cambridge, MA 02139, USA

FEATURES

source
 1..918
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 38..>918
 /gene="UGT1A1"
 /product="UDP-glucuronosyltransferase"
 <38..>918
 /gene="UGT1A1"
 /note="22..918; 1A1 variant allele; L233R"
 /allele="UGT1A1*33"
 38..>918
 /gene="UGT1A1"
 /codon_start=-1
 /product="UDP-glucuronosyltransferase"
 /protein_id="AAC4197.1"
 /db_xref="GI:12002135"
 /translation="MAVESQGRPLVGLLTVLPVYSHACKILLIPVDSHMLSL
 GAIQLOQRHEIVLAPASLYIDG"

CDS

QY 1 ttgtctctgcagaggtt 19
 |||||
 Db 323 TTTCCTCTGCCAGAGGTT 305

```
Query Match      100.0%; Score 19; DB 91; Length 3341
Best Local Similarity 100.0%; Pred. No. 13;
```

Wed Jul 25 10:29:08 2001

us-09-142-095-4.rge

Page 8

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgctcccgccagaggtt 19
|||||
Db 3185 ttgctcccgccagaggtt 3167

Search completed: July 25, 2001, 05:17:10
Job time: 9233 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:23:05 : Search time 247.87 Seconds
(without alignments)
48.131 Million cell updates/sec

Title: US-09-142-095-4

Perfect score: 19
Sequence: 1 ttgtctctgcagaggtc 19

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_0601.*
1: /SIDSI/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	19	100.0	19	AA79544
2	19	100.0	18	AA79540
3	18	100.0	620	AA79540
4	16	84.2	1167	AA033024
5	16	84.2	17	AA245058
6	16	84.2	17	AA245074
7	15.8	83.2	1307	AA795200
8	15.8	83.2	457	AA215518
9	15.8	83.2	843	AAV23955
10	15.8	83.2	1095	AAV68666
11	15.8	83.2	1156	AAV31924
			1571	AA27725

C 12	15.8	83.2	1945	21	AA098907	Human pancreatic c
C 13	15.8	83.2	2003	19	AAV22956	Nucleotide sequenc
C 14	15.8	83.2	2818	17	AA10554	Sheep mammary gland
C 15	15.8	83.2	3346	21	AA047622	Arabidopsis thaliana
C 16	14.8	77.9	239	21	AA10976	Fusarium venenatum
C 17	14.8	77.9	359	21	AA057066	Pinus radiata tran
C 18	14.8	77.9	375	21	AA057038	Pinus radiata tran
C 19	14.8	77.9	386	21	AA001025	Human secreted pro
C 20	14.8	77.9	404	21	AA003243	Human secreted pro
C 21	14.8	77.9	405	20	AA40504	Human secreted pro
C 22	14.8	77.9	430	21	AA057046	Pinus radiata tran
C 23	14.8	77.9	469	21	AA057049	Pinus radiata tran
C 24	14.8	77.9	471	21	AA056976	Partial cDNA clone
C 25	14.8	77.9	479	18	AA08072	Human secreted pro
C 26	14.8	77.9	479	18	AA02162	Human colon cancer
C 27	14.8	77.9	508	21	AA080032	Human breast cancer
C 28	14.8	77.9	530	21	AA057279	Human prostate can
C 29	14.8	77.9	697	20	AA033534	Human secreted pro
C 30	14.8	77.9	800	21	AA098339	Nucleotide sequenc
C 31	14.8	77.9	1485	19	AA198613	Nucleotide sequenc
C 32	14.8	77.9	1491	19	AA198614	Human angiotensin
C 33	14.8	77.9	2269	21	AA067774	Human angiotensin
C 34	14.8	77.9	2269	21	AA092213	Human angiotensin
C 35	14.8	77.9	2269	21	AA092213	Human angiotensin
C 36	14.8	77.9	2282	17	AA144321	Human TIE-2 ligand
C 37	14.8	77.9	2282	17	AA144321	Human TIE-2 ligand
C 38	14.8	77.9	2382	19	AA14650	Human TIE-2 ligand
C 39	14.8	77.9	3986	19	AAV18619	Human TIE-2 ligand
C 40	14.8	77.9	4500	21	AA087998	Candida albicans C
C 41	14.8	77.9	4590	21	AA075767	Human ORF1322
C 42	14.8	77.9	4594	20	AA077790	Cobra CYP1 coding
C 43	14.8	77.9	4548	20	AA085766	cDNA encoding cohr
C 44	14.8	77.9	45613	22	AA085535	Genomic fragment #
C 45	14.8	77.9	143068	21	AA021105	Human low adenosin

ALIGNMENTS

RESULT 1
AAT79544
ID AAT79544 standard; DNA; 19 BP.
AC AAT79544;
XX
DT 23-JAN-1998 (first entry)
XX
DE UGT1*1 gene exon 1 upstream PCR primer D.
XX
KM Uridine diphosphate glucuronosyltransferase gene; UGT;
KM Gilbert's syndrome; GS; unconjugated hyperbilirubinemia;
KM bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;
KM Drug trial efficiency; screening; PCR primer; ss.
XX
XX Synthetic.
OS Homo sapiens.
XX
PN M09732042-A2.
XX
XX 04-SEP-1997.
XX
PF 03-MAR-1997; 97NO-GB00577.
XX
PR 16-MAR-1996; 96GB-0005598.
PR 01-MAR-1996; 96GB-0004480.
XX
PA (UYDU-) UNIV DUNDIE.
XX
PI Burchell B;
XX
DR WPI; 1997-448702/41.
XX
PT Improving drug trial efficiency comprises identifying participants

PT with Gilbert's syndrome - useful as their altered drug metabolism
 PT may hinder result interpretation

PS Claim 14, Page 12, 31pp: English.

CC This PCR primer (with primers AAT79542-44) flanks the TATA box sequence
 CC upstream of the uridine diphosphate glucuronosyltransferase (UGT) gene
 CC 1*1 exon 1 (see AAT79540), and was used to amplify fragments of 98-100
 CC bp. This gene is known to be associated with Gilbert's syndrome (GS). GS
 CC is a mild, common form of unconjugated hyperbilirubinaemia associated
 CC with reduced bilirubin glucuronidation capacity. Analysis of the genetic
 CC basis of GS has allowed 2 forms to be identified. One is a mild form
 CC associated with a homozygous 2 bp insertion in the TATA sequence
 CC upstream of the UGT*1 exon 1, and the other is a more severe form
 CC associated with heterozygosity for a mutation which, when homozygous,
 CC causes Crigler-Najjar type 2 disease. The first form is autosomal
 CC recessive and the second is inherited dominantly. Patients suffering from
 CC GS, which is benign, may have altered metabolism of some drugs, making it
 CC difficult to determine if an effect is due to the drug or the syndrome.
 CC Drug trial efficiency would be improved if potential participants can be
 CC screened for the genetic basis of GS, and eliminated or included on
 CC basis of them possessing or not possessing GS.

CC Sequence 19 BP; 2 A; 5 C; 5 G; 7 T; 0 other;

Query Match 100.0%; Score 19; DB 18; Length 19;

Best Local Similarity 100.0%; Pred. No. 1.5; Mismatches 0; Gaps 0;

Matches 19; Conservative 0; Indels 0; Gaps 0;

OY 1 ttgtcctcgcagaggtt 19
 |||||
 DB 1 ttgtcctcgcagaggtt 19

RESULT 2

AAT79540/c

ID AAT79540 standard; DNA; 620 BP.

XX AAT79540:

XX 23-JAN-1998 (first entry)

DE Upstream DNA sequence of UGT*1 gene exon 1.

XX uridine diphosphate glucuronosyltransferase gene; UGT;

XX Gilbert's syndrome; GS; unconjugated hyperbilirubinaemia;

XX bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;

XX Drug trial efficiency; screening; ss.

OS Homo sapiens.

XX Location/Qualifiers

XX Key

XX protein_bind

XX GC_signal

XX misc-feature

XX protein_bind

XX protein_bind

XX protein_bind

XX misc-feature

XX misc-feature

XX misc-feature

XX misc-feature

XX misc-feature

XX misc-feature

FT protein_bind 386..392

FT /tag- g

FT /bound_moelty- AP1

FT /tag- h

FT /bound_moelty- AP1

FT /tag- 1

FT /note- "feature indicated in patent, but no further

FT explanation is given"

FT protein_bind

FT /tag- 1

FT /bound_moelty- HNF1

FT /tag- K

FT /note- "corresponds to positions -53 to -39 in patent"

FT /tag- 1

FT /note- "feature indicated in patent, but no further

FT explanation is given"

FT TATA_signal

FT /tag- m

FT /codon_start- 612

FT CDS

FT /tag- m

FT /codon_start- 612

FT /tag- m

FT /codon_start- 612

FT /tag- m

FT /codon_start- 612

FT /tag- m

FT /codon_start- 612

FT /tag- m

FT /codon_start- 612

FT /tag- m

FT /codon_start- 612

FT /tag- m

FT /codon_start- 612

FT /tag- m

FT /codon_start- 612

FT /tag- m

FT /codon_start- 612

FT /tag- m

FT /codon_start- 612

FT /tag- m

D8 605 TTGTGCTCCTGCCAGGGT 587

RESULT
ID AAQ33024/C standard; DNA; 1167 BP.

XX AC AAQ33024;
XX DT 27-JAN-1993 (first entry)
XX DE UGT1A Exon 1 from the UGT1 gene locus.
XX XX
XX UGT1A: UGT1BP; UGTIC: UGTID; UGTIE: UGTIF; Isozyme; bilirubin;
KM UDP-glucuronosyl transferase; CN; SS.
OS Homo sapiens.
XX Key Location/Oualifiers
FT misc_RNA 1..20 /tag-a "representation of 11.7 kbp of non-sequenced DNA between the sequences represented in AAQ33023 and AAQ33024"
FT FT /note-
FT TATA_signal 52..66 /*tag-b
FT exon 166..909 /*tag-c
FT /*label=UGT1A_Exon_1
FT mlsc_RNA 181..909 /*tag-d
FT /*tag-e "representation of 5 kbp of non-sequenced DNA between the sequences represented in AAQ33024 and AAQ33025"

PN WO9212987-A.
PD 06-AUG-1992.
PF 10-JAN-1992; 92MO-USO0282.
PR 10-JAN-1991; 91US-O639453.
PA (USSH) US DEP't HEALTH & HUMAN SERVICE.
PI Owens IS, Ritter JK.
DR WPI; 1992-284593/34.
P-PDBE; AAR30194.

Isolated gene locus UGT1, DNA segments and diagnostic probes - for diagnosing Gilbert's disease and Crigler-Najjar syndrome types I and II

Disclosure; Fig 1F; 9pp; English.

The isolated gene locus, UGT1, has a sequence of about 1000 bp which represent (1) Exon 1, comprising 6 transcriptional units (UTR1, E, D, C, BP and A), represented in AAQ27368 and AAQ33020-24 respectively;

(2) Exon 2, represented in AAQ33025;
(3) Exon 3, represented in AAQ33026;
(4) Exon 4, represented in AAQ33026;
(5) Exon 5, represented in AAQ33027; and
(6) about 69 kb of non-sequenced DNA.

Six unique N-terminal of 286-289 amino acids are encoded by the six different first exons and identical C-terminal of 246 amino acids are encoded by the common exons 2-5. The ugt1 gene locus encodes a family of UDP-glucuronosyl transferase isoymes, two of which metabolise bilirubin.

Sequence 1167 BP; 255 A; 259 C; 272 G; 340 T; 41 other:

Query Match 94.7%; Score 18; DB 13; Length 1167;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

2 ttgctctgcagagggt 19
|||||
174 TTGCTCTGTCGAGAGGTT 157

RESULT 4
AAZ45058/c
ID AAZ45058 standard: DNA; 17 BP.
AAZ45058:
28-FEB-2000 (first entry)

Forward PCR primer used in the secondary amplification of UGT1 exon 1A.

uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
unconjugated hyperbilirubinemia; drug metabolism; transgenic animal;
pharmacogenetic screening; diagnosis; PCR primer; ss.

Synthetic.
Homo sapiens.
MO9957322-42.
11-NOV-1999.
04-MAY-1999: 99MO-US09702.
07-MAY-1998: 98US-0084807.
(AXYS-) AXYS PHARM INC.
Penny L, Galvin M;
WPI: 2000-052981/04.

New nucleic acid representing polymorphisms in the human uridine
diphosphate glucuronosyltransferase gene, used for diagnosis and
evaluation of drug metabolism

Examples: Page 16; 63pp; English.

PCR primers AAZ45042-245073 are used to amplify human uridine
diphosphate-glucuronosyltransferase 1 (UGT1) exon sequences. The UGTs
are a family of enzymes that catalyse the glucuronic acid conjugation of
a wide range of endogenous and exogenous substrates including phenols,
alcohols, amines and fatty acids. Many of the reactions catalysed by
UGTs result in toxic substances being converted to compounds which are
more water soluble and are excreted. The invention relates to and
identifies UGT1 polymorphisms (AAZ45004-245041). The polymorphism
sequences are useful as probes for detecting UGT1 locus polymorphisms,
indicative of altered UGT1 expression or activity. These polymorphisms
are associated with Crigler-Najjar and Gilbert syndromes (unconjugated
hyperbilirubinemia) and drug metabolism. The genotyping of the UGT1 gene
is used to predict the rate of metabolism of UGT1 substrates, possible
drug-drug interactions and adverse side effects (i.e. to optimize drug
dosage), and to screen for diseases caused by exposure to toxins and to
study the effects of polymorphisms on enzymatic activity. The UGT1
sequences, including polymorphisms, can also be used to produce the
corresponding protein (or its fragments) or to generate transgenic
animals or modified cells e.g. for pharmacogenetic screening.

5Q Sequence 17 BP; 5 A; 4 C; 6 G; 2 T; 0 other;

Query Match 84.2%; Score 16; DB 21; Length 17;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ttgtctctgcagagc 16
|||||
DB 16 TTGTCTCTGCCAGAG 1

RESULT 5
AAZ45074/c

ID AAZ45074 standard; DNA; 17 BP.

AC AAZ45074;

DT 28-FEB-2000 (first entry)

DE Forward PCR primer for sequencing UGT1 exon 1A polymorphism #1.

XX uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;

KW glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;

KW unconjugated hyperbilirubinemia; drug metabolism; transgenic animal;

XX pharmacogenetic screening; diagnosis; PCR primer; ss.

OS Synthetic.

XX Homo sapiens.

PN MO9957322-A2.

PD 11-NOV-1999.

PF 04-MAY-1999; 99MO-US09702.

PR 07-MAY-1998; 98US-0084807.

XX (AXYS-) AXYS PHARM INC.

PI Penny L, Galvin M;

DR WPI: 2000-052981/04.

XX New nucleic acid representing polymorphisms in the human uridine

PT diphosphate glucuronosyltransferase gene, used for diagnosis and

PS evaluation of drug metabolism -

XX Examples: Page 19; 63pp; English.

CC Primers AAZ45074-245109 are used to sequence the human uridine

CC diphosphate-glucuronosyltransferase 1 (UGT1) exon polymorphism

CC sequences. The UGTs are a family of enzymes that catalyze the glucuronic

CC acid conjugation of a wide range of endogenous and exogenous substrates

CC including phenols, alcohols, amines and fatty acids. Many of the

CC reactions catalysed by UGTs result in toxic substances being converted

CC to compounds which are more water soluble and are excreted. The

CC invention relates to and identifies UGT1 polymorphisms (AAZ45004-245041).

CC The polymorphism sequences are useful as probes for detecting UGT1 locus

polymorphisms, indicative of altered UGT1 expression or activity. These

CC polymorphisms are associated with Crigler-Najjar and Gilbert syndromes

CC (unconjugated hyperbilirubinemia) and drug metabolism. The genotyping

CC of the UGT1 gene is used to predict the rate of metabolism of UGT1

CC substrates, possible drug-drug interactions and adverse side effects

CC (i.e. to optimize drug dosage), and to screen for diseases caused by

CC exposure to toxins and to study the effects of polymorphisms on

CC enzymatic activity. The UGT1 sequences, including polymorphisms, can also

CC be used to produce the corresponding protein (or its fragments) or to

CC generate transgenic animals or modified cells e.g. for pharmacogenetic

CC screening.

XX Sequence 17 BP; 5 A; 4 C; 6 G; 2 T; 0 other;

Query Match 84.2%; Score 16; DB 21; Length 17;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ttgtctctgcagagc 16
|||||
DB 16 TTGTCTCTGCCAGAG 1

RESULT 6
AAZ45200/c

ID AAZ45200 standard; DNA; 1307 BP.

AC AAZ45200;

DT 15-MAY-2001 (first entry)

DE Human 2hdelcalp DNA.

XX Notch receptor; ligand; cancer; melanoma; ischemia; ds.

XX Homo sapiens.

XX WO200112664-A2.

PD 22-FEB-2001.

PF 17-AUG-2000; 2000MO-US22609.

PR 19-AUG-1999; 99US-0149934.

XX (CHIR) CHIRON CORP.

PI Vivien C, Rohan M, Williams LT;

DR WPI: 2001-211201/21.

XX Novel notch receptor ligands useful for modulating angiogenesis and

PT immune responses for treating rheumatoid arthritis, cancer-related

PS angiogenesis to stop tumor growth, and as diagnostic reagents -

XX Disclosure: Fig 6; 66pp; English.

CC The present invention relates to a novel Notch receptor ligand.

CC The invention is useful for detecting Notch ligand expression in

CC human cancer cells or melanoma cells. Also useful for enhancing

CC angiogenesis in a mammal, useful when the mammal exhibits

CC tissue ischemia.

XX Sequence 1307 BP; 294 A; 328 C; 396 G; 289 T; 0 other;

Query Match 84.2%; Score 16; DB 22; Length 1307;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gctctgcagagc 19
|||||
DB 1119 GCCTCTGCCAGAGTT 1104

RESULT 7
AAZ16518/c

ID AAZ16518 standard; cDNA; 457 BP.

AC AAZ16518;

DT 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO: 3988.

XX Human; gene; gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;
 XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

XX MO9938972-A2.

XX 05-AUG-1999.

XX 28-JAN-1999; 99MO-US01619.

XX 03-APR-1998; 98US-0080666.

XX 24-FEB-1998; 98US-0072910.

XX 31-MAR-1998; 98US-0075954.

XX 03-APR-1998; 98US-0080114.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Cervenjakov R, Dickson M, Dermanac R, Dermanac S;
 XX Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 XX Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 XX Lamson G, Leshowitz D, Pol D, Randazzo F, Reinhard C;
 XX Stache-Crain B, Sudduth-Klinger J, Williams LP;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are
 XX differentially expressed in different cell types

XX Claim 1: Page 1892-1893; 2479pp; English.

XX The present invention describes a library of human polynucleotides
 XX comprising the sequences given in AA21532 to AA21779. Also described is
 XX a method of detecting differentially expressed genes correlated with the
 XX cancerous state of a mammalian cell, comprising detecting at least one
 XX differentially expressed gene product in a test sample from a cell
 XX suspected of being cancerous, where the gene product is encoded by one
 XX of the 5248 polynucleotide sequences given in AA21532 to AA21779. The
 XX polynucleotides can be used as a source of primers and probes, which can
 XX be used for a variety of purpose, e.g. detection of expression levels,
 XX mapping, tissue typing or profiling, forensics, genetic analysis and
 XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
 XX can be used for raising antibodies for experimental, diagnostic and
 XX therapeutic purposes. The polynucleotides may also be used to construct
 XX arrays for diagnostics (which may be used to determine function of an
 XX encoded protein); and to detect differences in expression levels between
 XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
 XX identify a genetic predisposition or susceptibility to a disease such as
 XX cancer). The polynucleotides of the invention are especially used in the
 XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
 XX and lung cancer. The polynucleotides can also be used to screen for
 XX peptide analogues and antagonists.

XX Sequence 457 BP; 120 A; 88 C; 93 G; 139 T; 17 other;

XX Query Match 83.2%; Score 15.8; DB 20; Length 457;

XX Best Local Similarity 89.5%; Pred. No. 83;

XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

XX 1 ttgtctctgcagaggtt 19

XX 300 TTGTCTCTGCTCCGAGAGCTT 282

XX RESULT 8

XX AA21532/c

XX ID AA21535 standard; CDNA; 843 BP.

XX AC AA21535;

XX XX

DT 04-AUG-1998 (first entry)

XX CDNA encoding human bone morphogenetic protein-16 (BMP-16).

XX Human; bone morphogenetic protein-16; BMP-16; murine protein; nodal;
 XX formation; bone; cartilage; treatment; wound healing; reduction;
 XX fibrosis; scar tissue formation; ss.

XX Homo sapiens.

XX Key

XX CDS Location/Qualifiers

XX sig_peptide

XX mat_peptide

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

DT 04-AUG-1998 (first entry)

XX CDNA encoding human bone morphogenetic protein-16 (BMP-16).

XX Human; bone morphogenetic protein-16; BMP-16; murine protein; nodal;
 XX formation; bone; cartilage; treatment; wound healing; reduction;
 XX fibrosis; scar tissue formation; ss.

XX Homo sapiens.

XX Key

XX CDS Location/Qualifiers

XX sig_peptide

XX mat_peptide

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

DT 04-AUG-1998 (first entry)

XX CDNA encoding human bone morphogenetic protein-16 (BMP-16).

XX Human; bone morphogenetic protein-16; BMP-16; murine protein; nodal;
 XX formation; bone; cartilage; treatment; wound healing; reduction;
 XX fibrosis; scar tissue formation; ss.

XX Homo sapiens.

XX Key

XX CDS Location/Qualifiers

XX sig_peptide

XX mat_peptide

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

DT 04-AUG-1998 (first entry)

XX CDNA encoding human bone morphogenetic protein-16 (BMP-16).

XX Human; bone morphogenetic protein-16; BMP-16; murine protein; nodal;
 XX formation; bone; cartilage; treatment; wound healing; reduction;
 XX fibrosis; scar tissue formation; ss.

XX Homo sapiens.

XX Key

XX CDS Location/Qualifiers

XX sig_peptide

XX mat_peptide

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

DT 04-AUG-1998 (first entry)

XX CDNA encoding human bone morphogenetic protein-16 (BMP-16).

XX Human; bone morphogenetic protein-16; BMP-16; murine protein; nodal;
 XX formation; bone; cartilage; treatment; wound healing; reduction;
 XX fibrosis; scar tissue formation; ss.

XX Homo sapiens.

XX Key

XX CDS Location/Qualifiers

XX sig_peptide

XX mat_peptide

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

DT 04-AUG-1998 (first entry)

XX CDNA encoding human bone morphogenetic protein-16 (BMP-16).

XX Human; bone morphogenetic protein-16; BMP-16; murine protein; nodal;
 XX formation; bone; cartilage; treatment; wound healing; reduction;
 XX fibrosis; scar tissue formation; ss.

XX Homo sapiens.

XX Key

XX CDS Location/Qualifiers

XX sig_peptide

XX mat_peptide

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

DT 04-AUG-1998 (first entry)

XX CDNA encoding human bone morphogenetic protein-16 (BMP-16).

XX Human; bone morphogenetic protein-16; BMP-16; murine protein; nodal;
 XX formation; bone; cartilage; treatment; wound healing; reduction;
 XX fibrosis; scar tissue formation; ss.

XX Homo sapiens.

XX Key

XX CDS Location/Qualifiers

XX sig_peptide

XX mat_peptide

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```

DT      18-JUN-1999      (first entry)
DE
XX      Human nodal protein encoding DNA.
KW      Nodal protein; lefty protein; TGF-beta; sexual development; human; bone;
KW      pituitary; cartilage; osteoarthritis; osteoporosis; haematopoiesis;
KW      periodontal disease; wound healing; tissue repair; tumour; cancer;
KW      interstitial lung disease; autoimmunity; leukaemia; lymphoma; immunity;
KW      immunosuppression; inflammatory bowel disease; myelosuppression;
KW      infectious disease; ss.
XX
OS      Homo sapiens.
XX
FH      Key                      Location/Qualifiers
FT      CDS                      1..852
FT                                     /tag= a
FT                                     /product= "Nodal protein"
FT                                     /note= "the start codon is not indicated"
FT      misc-feature             517..849
FT                                     /tag= b
FT                                     /note= "sequence coding for the active fragment of
FT                                     the Nodal polypeptide"
XX
PN      W09909198-A1.
XX
PD      25-FEB-1999.
XX
PF      20-AUG-1998;    98WO-0517211.
XX
PR      21-AUG-1997;    97US-0056565.
XX
PA      (HOMA-) HUMAN GENOME SCI INC.
XX
PI      Ebner R, Ruben SM, Soppet DR;
XX      WPI: 1999-190173/16.
XX      P-PSDB: AAI03849.
XX
XX      New isolate human Nodal and Lefty polypeptides
XX      Claim 2; Fig 1A; 182pp; English.
XX
XX      The present invention relates to novel human nodal and lefty proteins
XX      which are members of the TGF-beta family. The human nodal and lefty
XX      proteins may be involved in a developmental process such as the correct
XX      formation of various structures or in one or more post-developmental
XX      capacities including sexual development, pituitary hormone production,
XX      and the creation of bone and cartilage. The Nodal and lefty polypeptides
XX      are useful for enhancing or enriching the growth and/or differentiation
XX      of specific cell populations, eg. embryonic cells or stem cells. They can
XX      be used to treat such conditions as osteoarthritis, osteoporosis, and
XX      other abnormalities of bone, cartilage, muscle, tendon, ligament, and/or
XX      other connective tissues and/or organs such as liver, lung, cardiac,
XX      pancreas, and kidney. Compositions containing nodal and lefty proteins
XX      may be useful for growth formation, for treating periodontal disease and
XX      for modulating hematopoiesis, wound healing and tissue repair. They can
XX      also be used for the treatment of tumours, cancers, interstitial lung
XX      disease, and any deregulation of the growth and differentiation patterns
XX      of cell function including autoimmunity, arthritis, leukaemia, lymphomas,
XX      immunosuppression, immunity, humoral immunity, inflammatory bowel
XX      disease, myelosuppression, or infectious diseases. The present sequence
XX      represents a DNA encoding a human nodal polypeptide. The cDNA encoding
XX      the nodal protein is deposited under the ATCC deposit No. 209092 and/or
XX      209135.
XX
SO      Sequence 1156 bp; 285 A; 298 C; 340 G; 233 T; 0 other;

```

Db 421 TCTCTCTCCGACAGGCTT 403

RESULT 11

AA27725/C ID AAF1725 standard; CDNA: 1571 BP.

XX AAF77725;

DT 28-MAR-2001 (first entry)

XX Human transport protein TPT-25 coding sequence.

XX Human transport protein TPT-25 coding sequence.

XX Human transport protein TPT-25 coding sequence.

XX Human transport protein TPT-25 coding sequence.

XX Homo sapiens.

XX MO200078953-82.

XX 28-DEC-2000.

XX 16-JUN-2000; 2000MO-US16668.

XX 17-JUN-1999; 99US-013923.

XX 10-AUG-1999; 99US-014817.

XX 18-AUG-1999; 99US-0149357.

XX 28-OCT-1999; 99US-0162287.

XX (INCYTE) INCYTE GENOMICS INC.

XX [La] P, Yung J, Yue H, Hillman JL, Tang YT, Sandman G, Burford N;

XX Baughn MR, Alexander Y, Lu RM, Au-Yang J, Patterson C;

XX WPI: 2001-041424/05.

XX P-PSDB: AAB60105.

XX Isolated polypeptide with a human transport protein sequence is useful

XX for the diagnosis, prevention and treatment of disorders associated

XX with the immune, reproductive, and cardiovascular systems.

XX Claim 5: Page 155-156; 16pp; English.

XX The present invention provides the protein and coding sequences for 43

XX novel human transport proteins (designated TPTs). These can be used in

XX the diagnosis and treatment of transport, metabolic, neurological,

XX reproductive, cardiovascular, and other disorders, and cell proliferative

XX disorders such as cancer.

XX Sequence 1571 BP; 572 A; 265 C; 289 G; 495 T; 0 other:

XX Query Match 83.2% Score 15.8; DB 22; Length 1571;

XX Best Local Similarity 89.5% Pred. No. 96;

XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

XX Db 153 TCTCTCTCCGACAGGCTT 1516

XX RESULT 12

XX AAC98907/C

XX AAC98907 standard; CDNA: 1945 BP.

XX AAC98907;

XX 09-MAR-2001 (first entry)

XX Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:135.

XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;

XX detection; diagnosis; identification; cytostatic; neuroprotective;

XX neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;

XX anti-inflammatory; cardiac; gene therapy; chromosome mapping; linkage

XX analysis; tissue identification; tissue typing; forensic;

XX neutral; immune system; muscular; reproductive; gastrointestinal;

XX pulmonary; cardiovascular; renal; proliferative; ss.

XX Homo sapiens.

XX MO200055320-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000MO-US05989.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENE SCI INC.

XX Rosen CA, Ruben SM;

XX WPI: 2000-579444/54.

XX P-PSDB: AAB54142.

XX New nucleic acid that is a pancreatic cancer antigen for preventing,

XX treating, or ameliorating a medical condition, particular pancreatic

XX cancer, or for use in assays for diagnosing a pathological condition.

XX Claim 1: Page 594-595; 1379pp; English.

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated

XX proteins, called pancreatic cancer antigens, given in AAB54008 to

XX AAB54466. The human pancreatic cancer antigens have cytostatic,

XX neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,

XX gynaecological, cardiac and anti-inflammatory activities, and can be used

XX in gene therapy. The polynucleotide and proteins can be used for

XX preventing, treating, or ameliorating a medical condition or in assays

XX for diagnosing a pathological condition or a susceptibility to one in a

XX subject. Binding partners to the proteins and the activity of the

XX proteins can be identified. The pancreatic cancer antigens can be used to

XX detect, treat or prevent pancreatic disorders, especially cancer.

XX Agonists and antagonists to the antigens can be screened for. The

XX pancreatic cancer antigen polynucleotide can be used to design nucleic

XX acid hybridization probes that can be used in chromosome mapping, linkage

XX analysis, tissue identification and/or typing and a variety of forensic

XX and diagnostic methods. The proteins can be used to generate antibodies

XX which are used to purify, detect and target the polypeptides, including

XX both in vivo and in vitro diagnostic and therapeutic methods. The

XX proteins can be used to treat, prevent or prevent neural, immune system, muscular,

XX reproductive, gastrointestinal, pulmonary, cardiovascular, renal or

XX proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent

XX sequences used in the exemplification of the present invention.

XX Sequence 1945 BP; 533 A; 300 C; 477 G; 552 T; 13 other:

XX Query Match 83.2% Score 15.8; DB 21; Length 1945;

XX Best Local Similarity 89.5% Pred. No. 98;

XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

XX Db 1203 TTTACTCTCCGACAGCTT 1185

XX RESULT 13

XX AAV22956/C

XX AAV22956 standard; DNA: 2003 BP.

XX AAV22956;

XX 04-AUG-1998 (first entry)

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134283.
PR 14-MAY-1999; 99US-0134286.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 26-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142554.
PR 06-JUL-1999; 99US-0142955.
PR 08-JUL-1999; 99US-0142990.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144081.
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145819.
PR 28-JUL-1999; 99US-0145851.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0150820.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154038.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156599.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158233.

XX Nucleotide sequence of the second exon of human BMP-16.
 DE Human; bone morphogenetic protein-16; BMP-16; murine protein; nodal;
 XX formation; bone; cartilage; treatment; wound healing; reduction;
 KW fibrosis; scar tissue formation; ss.
 XX Homo sapiens.
 OS W09812322-AL.
 XX 26-MAR-1998.
 XX 09-JUL-1997; 97MO-US11954.
 XX 18-SEP-1996; 96US-0715202.
 XX (GENY) GENETICS INST INC.
 XX Celeste AJ, Murray BL;
 PI WPI; 1998-217262/19.
 DR New isolated bone morphogenetic protein-16 - used to develop
 XX products for inducing formation of bone, cartilage and other
 PT connective tissue, particularly for wound healing and tissue repair
 PS Disclosure; Pages 36-37; 43pp; English.
 XX The present sequence represents the nucleotide sequence of the second
 CC exon of the genomic DNA for human bone morphogenetic protein-16 (BMP-16).
 CC Human BMP-16 is a homologue of a murine protein called nodal, which is
 CC expressed in the mouse node during gastrulation. The BMP-16 proteins can
 CC induce the formation of bone, cartilage or other connective tissue. They
 CC can be used for treating bone, cartilage or other connective tissue
 CC defects, periodontal disease or healing of various types of tissues and
 CC wounds. They can also increase neuronal, astrocytic and glial cell
 CC survival and therefore be useful in transplantation and treatment of
 CC conditions exhibiting a decrease in neuronal survival and repair. They
 CC can also exhibit properties such as angiogenic, chemotactic and/or
 CC chemotactant properties, and effects on cells including induction of
 CC collagen synthesis, fibrosis, differentiation responses, cell
 CC proliferative responses and responses involving cell adhesion, migration
 CC and extracellular matrices. These properties make the proteins potential
 CC agents for wound healing, reduction of fibrosis and reduction of scar
 CC tissue formation.
 CC
 CC Sequence 2003 BP; 471 A; 490 C; 570 G; 471 T; 1 other;
 SQ

Query Match 83.2%; Score 15.8; DB 19; Length 2003;
 Best Local Similarity 89.5%; Pred. No. 99;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttgtctctgcagaggtt 19
 | ||||| |||||
 DB 802 TCTGCTCTGCGAGAGGT 784

RESULT 14
 AAT10554
 ID AAT10554 standard; cDNA to mRNA; 2818 BP.
 XX AAT10554;
 AC
 XX 03-APR-1996 (first entry)
 DE Sheep mammary gland factor cDNA.
 XX Mammary gland factor; signal transduction; lactogenic hormone;
 KW cytokine regulated transcription factor; transgenic animal; sheep; ss.
 XX Ovis aries.
 OS

XX Key Location/Qualifiers
 FH CDS 244..2628
 FT /tag= a
 XX CA2139384-A.
 XX 12-NOV-1995.
 XX 30-DEC-1994; 94CA-2139384.
 XX 11-MAR-1994; 94GB-0009396.
 XX (CIBA) CIBA GEIGY AG.
 XX Gouilleux F, Groner B, Makao H;
 PI WPI; 1996-077786/09.
 DR P-PSDB; AAR8199.
 XX DNA encoding mammary gland factor protein - used to identify cpds.
 PT affecting intracellular signal transduction of a lactogenic hormone,
 PT or cytokine
 PS Claim 2; Page 28-34; 42pp; English.
 XX A cDNA clone (AAT10554) coding for mammary gland factor (MGF)
 CC (AAR8199) was obtd. from a cDNA library derived from sheep
 CC lactating mammary tissue mRNA using probes (AAT1055-56) based on
 CC internal peptides of MGF. The cDNA is used to produce recombinant
 CC MGF in host cells, or as a probe. Transgenic animals, e.g. sheep,
 CC overexpressing MGF can be produced in order to increase milk prodn.
 CC or to produce a therapeutically useful protein.
 CC
 CC Sequence 2818 BP; 618 A; 861 C; 819 G; 520 T; 0 other;
 SQ

Query Match 83.2%; Score 15.8; DB 17; Length 2818;
 Best Local Similarity 89.5%; Pred. No. 1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttgtctctgcagaggtt 19
 | ||||| |||||
 DB 2599 ctactctgcagaggtc 2617

RESULT 15
 AAC47622
 ID AAC47622 standard; DNA; 3346 BP.
 XX AAC47622;
 AC
 XX 18-OCT-2000 (first entry)
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54504.
 XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX Arabidopsis thaliana.
 OS
 XX EP1033405-A2.
 XX 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 XX 05-MAR-1999; 99US-0123180.
 XX 03-MAR-1999; 99US-0123548.
 XX 23-MAR-1999; 99US-0125788.
 XX 25-MAR-1999; 99US-0126264.

PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 83.2%; Score 15.8; DB 21; Length 3346;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttgtctctgcagaggtt 19
|||||
Db 1742 ttgtctctgcagaggtt 1760

Search completed: July 25, 2001, 05:23:06
Job time: 4679 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:18:53 / Search time 117.39 seconds

(without alignments)
29.985 Million cell updates/sec

Title: US-09-142-095-4

Sequence: 1 ttgtctctgcagaggtt 19

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents, NA:*

- 1: /cgn2_6/ptodata/2/line/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/line/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/line/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/line/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/line/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/line/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19	100.0	1190	5	PCT-US92-00282-18
C 2	15.8	83.2	843	2	US-08-715-202A-1
C 3	15.8	83.2	2002	1	US-08-715-202A-3
C 4	15.8	83.2	2818	1	US-08-366-276-1
C 5	14.8	77.9	2282	1	US-08-373-579-5
C 6	14.8	77.9	2282	1	US-08-418-595-5
C 7	14.8	77.9	2282	1	US-08-665-926-5
C 8	14.8	77.9	2282	4	US-09-162-437-5
C 9	14.8	77.9	5924	2	US-08-447-411-44
C 10	14.8	77.9	5948	1	US-08-662-227-1
C 11	14.4	75.8	1858	2	US-08-359-705B-7
C 12	14.4	75.8	1858	2	US-08-286-846A-7
C 13	14.4	75.8	1858	2	US-08-457-880A-7
C 14	14.4	75.8	1858	3	US-08-444-622A-7
C 15	14.4	75.8	1858	3	US-08-942-562-7
C 16	14.4	75.8	1858	4	US-09-136-923-7
C 17	14.4	75.8	2715	2	US-08-359-705B-5
C 18	14.4	75.8	2715	2	US-08-286-846A-5
C 19	14.4	75.8	2715	2	US-08-457-880A-5
C 20	14.4	75.8	2715	3	US-08-444-622A-5
C 21	14.4	75.8	2715	3	US-08-942-562-5
C 22	14.4	75.8	2715	4	US-09-136-923-5
C 23	14.4	75.8	2940	1	US-08-286-846A-8
C 24	14.4	75.8	2940	2	US-08-441-104A-8
C 25	14.4	75.8	2940	2	US-08-440-816A-8
C 26	14.2	74.7	300	2	US-07-938-154-6
C 27	14.2	74.7	300	5	PCT-US91-03311-6

28	14.2	74.7	693	1	US-08-052-205-10	Sequence 10, Appl
29	14.2	74.7	693	1	US-08-595-974-10	Sequence 10, Appl
30	14.2	74.7	759	1	US-08-052-205-8	Sequence 8, Appl1
31	14.2	74.7	759	1	US-08-595-974-8	Sequence 8, Appl1
32	14.2	74.7	1044	1	US-08-052-205-6	Sequence 6, Appl1
33	14.2	74.7	1044	1	US-08-595-974-6	Sequence 6, Appl1
34	14.2	74.7	1062	1	US-08-052-205-1	Sequence 1, Appl1
35	14.2	74.7	1062	1	US-08-595-974-1	Sequence 1, Appl1
36	14.2	74.7	1110	1	US-08-052-205-5	Sequence 5, Appl1
37	14.2	74.7	1110	1	US-08-595-974-5	Sequence 5, Appl1
38	14.2	74.7	1393	1	US-08-052-205-2	Sequence 2, Appl1
39	14.2	74.7	1393	1	US-08-595-974-2	Sequence 2, Appl1
40	14.2	74.7	1449	2	US-08-705-868-2	Sequence 2, Appl1
41	14.2	74.7	1449	3	US-09-123-635-2	Sequence 2, Appl1
42	14.2	74.7	1451	1	US-08-031-143B-68	Sequence 66, Appl
43	14.2	74.7	1470	1	US-08-052-205-3	Sequence 3, Appl1
44	14.2	74.7	1470	1	US-08-595-974-3	Sequence 3, Appl1
45	14.2	74.7	1608	2	US-08-424-224-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
PCT-US92-00282-18/c

Sequence 18, Application PC/TUS9200282

GENERAL INFORMATION:

APPLICANT: OMENS, IDA S.

APPLICANT: RITTER, JOSEPH K.

TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION THEREIN.

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESS: CUSHMAN DARBY & CUSHMAN

STREET: 1615 L STREET, N.W.

CITY: WASHINGTON

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20036-5601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/00282

FILING DATE: 19920110

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: SCOTT, WATSON T.

REGISTRATION NUMBER: 26581

REFERENCE/DOCKET NUMBER: 91532-PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-861-3000

TELEFAX: 202-822-0944

TELEX: 6714627 CUSH

INFORMATION FOR SEQ. ID NO. 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 1190 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

PCT-US92-00282-18

Query Match 100.0%; Score 19; DB 5; Length 1190;
Best Local Similarity 100.0%; Pred. No. 0.49;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 1 ttgtctctgcagaggtt 19
DB 78 ttgtctctgcagaggtt 60

REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: 4-19992/A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2818 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 244..2625
OTHER INFORMATION: /product= "mammary gland factor"

Query Match 83.2%; Score 15.8; DB 1; Length 2818;
Best Local Similarity 89.5%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ttgtctctgcagaggtt 19
Db 2599 TTTACTCTGCGCAGAGCT 2617

RESULT 5
US-08-373-579-5
Sequence 5, Application US/08373579
Patent No. 5650490
GENERAL INFORMATION:
APPLICANT: Davis, et al.
TITLE OF INVENTION: TIE-2 LIGAND, METHOD OF MAKING AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill River Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/373,579
FILING DATE: 17-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,503
FILING DATE: 09-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,492
FILING DATE: 02-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/330,261
FILING DATE: 27-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/319,932
FILING DATE: 07-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J.
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 330-D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 345-7400
TELEFAX: (914) 345-7721

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2282 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 357..1847
US-08-373-579-5

Query Match 77.9%; Score 14.8; DB 1; Length 2282;
Best Local Similarity 88.9%; Pred. No. 75;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ttgtctctgcagaggtt 18
Db 488 TTTCTCTGCGCAGAGAT 505

RESULT 6
US-08-418-595-5
Sequence 5, Application US/08418595
Patent No. 5814464
GENERAL INFORMATION:
APPLICANT: Davis, et al.
TITLE OF INVENTION: TIE-2 LIGAND, METHOD OF MAKING AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill River Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/418,595
FILING DATE: 06-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/373,579
FILING DATE: 17-JAN-1995
APPLICATION NUMBER: US 08/353,503
FILING DATE: 09-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,492
FILING DATE: 02-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/330,261
FILING DATE: 27-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/319,932
FILING DATE: 07-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J.
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 330-D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 345-7400
TELEFAX: (914) 345-7721
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2282 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 357..1847
US-08-418-595-5

Query Match 77.9%; Score 14.8; DB 1; Length 2282;
Best Local Similarity 88.9%; Pred. No. 75;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttgtctctgcagaggt 18
DB 488 ttctctctgcagagat 505

RESULT 7
US-08-665-926-5
Sequence 5, Application US/08665926
Patent No. 5851797
GENERAL INFORMATION:
APPLICANT: Valenzuela et al.
TITLE OF INVENTION: THE LIGAND-3, METHODS OF MAKING AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill River Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-6707
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,926
FILING DATE: 19-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Robert J. Cobert
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 330-H
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 345-7400
TELEFAX: (914) 345-2113
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2282 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 357..1847
US-08-665-926-5

Query Match 77.9%; Score 14.8; DB 2; Length 2282;
Best Local Similarity 88.9%; Pred. No. 75;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttgtctctgcagaggt 18
DB 488 ttctctctgcagagat 505

RESULT 8
US-09-162-437-5

Sequence 5, Application US/09162437
Patent No. 6166185

GENERAL INFORMATION:
APPLICANT: Davis, et al.
TITLE OF INVENTION: TIF-2 LIGAND, METHOD OF MAKING AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill River Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/162,437
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/418,595
FILING DATE: 06-APR-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/373,579
FILING DATE: 17-JAN-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,503
FILING DATE: 09-DEC-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,492
FILING DATE: 02-DEC-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/330,261
FILING DATE: 27-OCT-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/319,932
FILING DATE: 07-OCT-1994

ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J.
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 330-D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 345-7400
TELEFAX: (914) 345-7721

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2282 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 357..1847
US-09-162-437-5

Query Match 77.9%; Score 14.8; DB 4; Length 2282;
Best Local Similarity 88.9%; Pred. No. 75;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttgtctctgcagaggt 18
DB 488 ttctctctgcagagat 505

RESULT 9
US-08-447-411-44
Sequence 44, Application US/08447411
Patent No. 5773243

GENERAL INFORMATION:
APPLICANT: FRITZINGER, DAVID C.
APPLICANT: BREDEHORST, REINHARD
APPLICANT: VOGEL, CARL-WILHELM
TITLE OF INVENTION: DNA ENCODING COBRA C3, CVF1, AND CVF2
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,411
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/043,747
FILING DATE: 07-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, NO. 5773243man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 1126-101-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 24855 DPAT UR
INFORMATION FOR SEQ. ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 5924 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 4..69
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 70..4929
FEATURE:
NAME/KEY: CDS
LOCATION: 4..4929
US-08-447-411-44

Query Match 77.9%; Score 14.8; DB 1; Length 5924;
Best Local Similarity 88.9%; Pred. No. 87;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ttgtctctgcagaggt 18
|| ||||| |||||
Db 5567 TTGGCTCTGTGACAGAGGT 5584

RESULT 10
US-08-662-227-1
Sequence 1, Application US/08662227
Patent No. 592320
GENERAL INFORMATION:
APPLICANT: VOGEL, CARL-WILHELM
APPLICANT: BREDEHORST, REINHARD
APPLICANT: KOCK, MICHAEL
APPLICANT: FRITZINGER, DAVID
TITLE OF INVENTION: RECOMBINANT PROCVF
NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/662,227
FILING DATE: 14-JUN-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 1126-0107-0X
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5948 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-662-227-1

Query Match 77.9%; Score 14.8; DB 2; Length 5948;
Best Local Similarity 88.9%; Pred. No. 87;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ttgtctctgcagaggt 18
|| ||||| |||||
Db 5567 TTGGCTCTGTGACAGAGGT 5584

RESULT 11
US-08-359-705B-7/c
Sequence 7, Application US/08359705B
Patent No. 5844092
GENERAL INFORMATION:
APPLICANT: Presta, Leonard G.
APPLICANT: Shelton, David L.
APPLICANT: Ufer, Roman
TITLE OF INVENTION: Human tkr Receptors and Neurotrophic Factor Inhibitors
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinpatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/359,705B
FILING DATE: 20-Dec-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/286646
FILING DATE: 08/10/94
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/215139

FILING DATE: 03/18/94

ATTORNEY/AGENT INFORMATION:

NAME: TORCHIA, PHD., Timothy E.

REGISTRATION NUMBER: 36,700

REFERENCE/DOCKET NUMBER: P0873P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-8674

TELEFAX: 650/952-9861

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1858 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

US-08-359-7058-7

Query Match 75.8%; Score 14.4; DB 2; Length 1858;
Best Local Similarity 93.8%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgcctctgccagagct 18
|||||
DB 545 TGCTCCTGCCAGAGCT 530

RESULT 12

US-08-286-846A-7/C

Sequence 7, Application US/08286846A

Patent No. 5877016

GENERAL INFORMATION:

APPLICANT: Prestia, Leonard G.

APPLICANT: Shelton, David L.

TITLE OF INVENTION: Human trk Receptors and Neurotrophic Factor Inhibitors

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:

ADDRESSER: Genentech, Inc.

STREET: 460 Point San Bruno Blvd

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Winpatin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/286,846A

FILING DATE: 05-Aug-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: TORCHIA, PHD., Timothy E.

REGISTRATION NUMBER: 36,700

REFERENCE/DOCKET NUMBER: P0873P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-8674

TELEFAX: 415/952-9861

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1858 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

US-08-286-846A-7

Query Match 75.8%; Score 14.4; DB 2; Length 1858;
Best Local Similarity 93.8%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgcctctgccagagct 18
|||||
DB 545 TGCTCCTGCCAGAGCT 530

RESULT 13

US-08-457-880A-7/C

Sequence 7, Application US/08457880A

Patent No. 5910574

GENERAL INFORMATION:

APPLICANT: Leonard G. Prestia

APPLICANT: David L. Shelton

TITLE OF INVENTION: HUMAN trk RECEPTORS AND NEUROTROPHIC FACTOR

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:

ADDRESSER: Genentech, Inc.

STREET: 1 DNA Way

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Winpatin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/457,880A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/444,622

FILING DATE: 19-May-1995

APPLICATION NUMBER: 08/286846

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: TORCHIA, PHD., Timothy E.

REGISTRATION NUMBER: 36,700

REFERENCE/DOCKET NUMBER: P0873P1C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-8674

TELEFAX: 650/952-9861

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1858 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

US-08-457-880A-7

Query Match 75.8%; Score 14.4; DB 2; Length 1858;
Best Local Similarity 93.8%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgcctctgccagagct 18
|||||
DB 545 TGCTCCTGCCAGAGCT 530

RESULT 14

US-08-444-622A-7/C

Sequence 7, Application US/08444622A

Patent No. 6025166

GENERAL INFORMATION:

APPLICANT: Leonard G. Prestia

APPLICANT: David L. Shelton

TITLE OF INVENTION: HUMAN trk RECEPTORS AND NEUROTROPHIC FACTOR INHIBITORS

NUMBER OF SEQUENCES: 41

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Genentech, Inc.
;; STREET: 1 DNA Way
;; CITY: South San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94080
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Minipalin (Genentech)
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/444,622A
;; FILING DATE: 19-May-1995
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/286846
;; FILING DATE: 5
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Torchia, PhD., Timothy E.
;; REGISTRATION NUMBER: 36,700
;; REFERENCE/DOCKET NUMBER: P0873PIC3
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650/225-8674
;; TELEFAX: 650/952-9881
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1858 base pairs
;; TYPE: Nucleic Acid
;; STRANDEDNESS: Single
;; TOPOLOGY: Linear
;;
US-08-444-622A-7
;;
Query Match 75.8%; Score 14.4; DB 3; Length 1858;
Best Local Similarity 93.8%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 tgcctctgccagaggt 18
|||||
DB 545 TGCCTCTGCCAGAGCT 530
;;
RESULT 15
US-08-942-562-7/C
;; Sequence 7, Application US/08942562
;; Patent No. 6027927
;; GENERAL INFORMATION:
;; APPLICANT: Presta, Leonard G.
;; APPLICANT: Shelton, David L.
;; APPLICANT: Ulfert, Roman
;; TITLE OF INVENTION: Human trk Receptors and Neurotrophic
;; NUMBER OF SEQUENCES: 37
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Genentech, Inc.
;; STREET: 460 Point San Bruno Blvd
;; CITY: South San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94080
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Minipalin (Genentech)
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/942,562
;; FILING DATE: 01-OCT-1997
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/444,597

;; FILING DATE: 19-May-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Torchia, PhD., Timothy E.
;; REGISTRATION NUMBER: 36,700
;; REFERENCE/DOCKET NUMBER: P0873PIC2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415/225-8674
;; TELEFAX: 415/952-9881
;; TELEX: 910/371-7168
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1858 base pairs
;; TYPE: Nucleic Acid
;; STRANDEDNESS: Single
;; TOPOLOGY: Linear
;;
US-08-942-562-7
;;
Query Match 75.8%; Score 14.4; DB 3; Length 1858;
Best Local Similarity 93.8%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 tgcctctgccagaggt 18
|||||
DB 545 TGCCTCTGCCAGAGCT 530

Search completed: July 25, 2001, 05:18:55
Job time: 9218 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on:

July 25, 2001, 02:01:07 ; Search time 2762.24 Seconds
(without alignments)
71.866 Million cell updates/sec

Title:

US-09-142-095-1

Perfect score:

21

Sequence:

1 aagtgacctccctgtacctt 21

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched:

10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters:

20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_esthum29.*
64: em_esthum30.*
65: em_esthum31.*
66: em_esthum32.*
67: em_esthum33.*
68: em_esthum34.*
69: em_esthum35.*
70: em_esthum36.*
71: em_esthum37.*
72: em_esthum38.*
73: em_esthum39.*
74: em_esthum40.*
75: em_esthum41.*
76: em_esthum42.*
77: em_esthum43.*
78: em_esthum44.*
79: em_esthum45.*
80: em_esthum46.*
81: em_esthum47.*
82: em_esthum48.*
83: em_esthum49.*
84: em_esthum50.*
85: em_esthum51.*
86: em_esthum52.*
87: em_esthum53.*
88: em_esthum54.*
89: em_esthum55.*
90: em_esthum56.*
91: em_esthum57.*
92: em_esthum58.*
93: em_esthum59.*
94: em_esthum60.*
95: em_esthum61.*
96: em_esthum62.*
97: em_esthum63.*
98: em_esthum64.*
99: em_esthum65.*
100: em_esthum66.*
101: em_esthum67.*
102: em_esthum68.*
103: em_esthum69.*
104: em_esthum70.*
105: em_esthum71.*
106: em_esthum72.*
107: em_esthum73.*
108: em_esthum74.*
109: em_esthum75.*
110: em_esthum76.*
111: em_esthum77.*
112: em_esthum78.*
113: em_esthum79.*
114: em_esthum80.*
115: em_esthum81.*
116: em_esthum82.*

117: gb_esc18:*
118: gb_esc49:*
119: gb_esc50:*
120: gb_esc51:*
121: gb_esc52:*
122: gb_esc53:*
123: gb_esc54:*
124: gb_esc55:*
125: gb_esc56:*
126: gb_esc57:*
127: gb_esc58:*
128: gb_esc59:*
129: gb_esc60:*
130: gb_esc61:*
131: gb_esc62:*
132: gb_esc63:*
133: gb_esc64:*
134: gb_esc65:*
135: gb_esc66:*
136: gb_esc67:*
137: gb_esc68:*
138: gb_esc69:*
139: gb_esc70:*
140: gb_esc71:*
141: gb_esc72:*
142: gb_esc73:*
143: gb_esc74:*
144: gb_esc75:*
145: gb_esc76:*
146: gb_esc77:*
147: gb_esc78:*
148: gb_esc79:*
149: gb_esc80:*
150: gb_esc81:*
151: gb_esc82:*
152: gb_esc83:*
153: gb_esc84:*
154: gb_esc85:*
155: gb_esc86:*
156: gb_esc87:*
157: gb_esc88:*
158: gb_esc89:*
159: gb_esc90:*
160: gb_esc91:*
161: gb_esc92:*
162: gb_esc93:*
163: gb_esc94:*
164: gb_esc95:*
165: gb_esc96:*
166: gb_esc97:*
167: gb_esc98:*
168: gb_esc99:*
169: gb_esc100:*
170: gb_esc101:*
171: gb_esc102:*
172: gb_esc103:*
173: gb_esc104:*
174: gb_esc105:*
175: gb_esc106:*
176: gb_esc107:*
177: gb_esc108:*
178: gb_esc109:*

190: gb_esc110:*
191: gb_esc111:*
192: gb_esc112:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.8	84.8	212	188	R60857 yj98e12.r1
2	17.8	84.8	348	187	R31201 yj62a02.r1
3	17.8	84.8	417	188	R67982 y104a10.r1
4	17.8	84.8	418	187	R31714 yj06a12.r1
5	17.8	84.8	481	188	R63154 y101a11.r1
6	17.8	84.8	498	225	A0882343 y5382.r1
7	17.8	84.8	510	225	A0185931 HS_-3065.r1
8	17.8	84.8	608	230	A0511067 RPCI-11.r1
9	17	81.0	471	31	AV553454 AV553454
10	16.8	80.0	193	115	AM381670 QVO-FHT03
11	16.8	80.0	304	8	AA500281 V197f08.r
12	16.8	80.0	326	9	AA560193 v120g01.r
13	16.8	80.0	410	9	AA620205 v064e06.r
14	16.8	80.0	416	22	A1619361 v197f08.y
15	16.8	80.0	439	9	AA616429 v069e02.r
16	16.8	80.0	469	8	AA473687 v989c05.x
17	16.8	80.0	508	13	AA881891 vx29b10.r
18	16.8	80.0	514	236	A0978201 RPCI-23.r
19	16.8	80.0	535	238	A2093040 RPCI-23.r
20	16.8	80.0	547	120	AW743288 up64h01.y
21	16.8	80.0	556	2	AA108719 mp29f06.y
22	16.8	80.0	561	238	A2123144 RPCI-23.r
23	16.8	80.0	740	245	A2616160 RPCI-23.r
24	16.4	78.1	322	251	A2902873 RPCI-24.r
25	16.4	78.1	415	147	BF370208 RCA-FM00345
26	16.4	78.1	481	114	AM442557 GtHEST411
27	16.4	78.1	508	21	A1499582 cn96e01.x
28	16.4	78.1	530	123	AM981162 PC18D09.f
29	16.4	78.1	665	240	A2636471 RPCI-23.r
30	16.4	78.1	668	246	A2617251 1M0448M22
31	16.4	78.1	783	155	RG573720 6025947885
32	16.4	78.1	810	221	CNS0303RY AL433367
33	16.4	78.1	911	147	BF304665 601888077
34	16.4	78.1	914	220	CNS0107C5 FL166530
35	16.4	78.1	1049	222	CNS0588X AF149685
36	16.2	77.1	193	218	AA149685 AA716221
37	16.2	77.1	265	11	AA716221 zh28b1.s
38	16.2	77.1	274	23	A1651653 A1021546
39	16.2	77.1	285	15	ub09c11.r
40	16.2	77.1	319	14	AA968674 cr71e07.s
41	16.2	77.1	340	24	A1744438 wtf9c03.x
42	16.2	77.1	351	4	AA235220 z536e12.x
43	16.2	77.1	355	123	AM999271 MRO-BN00703
44	16.2	77.1	370	18	A1274481 gv51d06.x
45	16.2	77.1	378	168	BF722600 mah21c09

ALIGNMENTS

```

TITLE
The WashU-Merck EST Project
JOURNAL
Unpublished (1995)
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 761
High quality sequence stops: 74
Source: IMAGE Consortium, LNLN
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 761 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 74.
Location/Qualifiers
1. 212
/organism="Homo sapiens"
/db_xref="GDB:544134"
/db_xref="taxon:9606"
/clone="IMAGE:137806"
/clone_lib="Scars Placenta Nb2hp"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pTR733 (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
stranded cDNA was primed with a Not I - oligo(dT) primer [5'
AAGTCGAGAATTCGCCGCACGCAAGAAATTTTCTTTTTTTTTTTTCTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTR733 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Donato."
BASE COUNT      50 a          63 c          32 g          66 t          1 others
ORIGIN
Query Match      84.8%; Score 17.8; DB 188; Length 212;
Best Local Similarity 90.5%; Pzed. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY      1 aaagtaactccctgctacctt 21
|||||
Db      11 AAGTAAGCTCCCTGCTACTCTT 31

FEATURES
source
1. 212
/organism="Homo sapiens"
/db_xref="GDB:544134"
/db_xref="taxon:9606"
/clone="IMAGE:137806"
/clone_lib="Scars Placenta Nb2hp"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pTR733 (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
stranded cDNA was primed with a Not I - oligo(dT) primer [5'
AAGTCGAGAATTCGCCGCACGCAAGAAATTTTCTTTTTTTTTTTTCTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTR733 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Donato."
BASE COUNT      50 a          63 c          32 g          66 t          1 others
ORIGIN
Query Match      84.8%; Score 17.8; DB 188; Length 212;
Best Local Similarity 90.5%; Pzed. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY      1 aaagtaactccctgctacctt 21
|||||
Db      11 AAGTAAGCTCCCTGCTACTCTT 31

RESULT      2
LOCUS      R31201      348 bp      mRNA      EST
28-APR-1995
```

Email: estevason.wustl.edu

Insert Size: 728

High quality sequence stops: 195

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 728

Seq primer: M13RPI

Std Error: 0.00

High quality sequence stop: 195.

Location/Qualifiers

source

1. 348

/organism="Homo sapiens"

/db_xref="GDB:540030"

/db_xref="taxon:9606"

/clone="IMAGE:134282"

/clone_lib="Soares placenta Nb2HP"

/sex="Female"

/dev_stage="Placenta obtained at birth (full term)"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: placenta; Vector: pT73D (Pharmacia) with a

modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

AACTGAGAGATTCGGCGCCGAGATTTTCTTTTCTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization. Library

constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT

78 a 92 c 66 g 108 t 4 others

Db 10 AAGTAGCTCCCTGCTACTT 30

Query Match 84.8%; Score 17.8; DB 187; Length 348;

Best Local Similarity 90.5%; Pred. No. 1.5e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aagtgactccctgactct 21

||||| |||||||

Db 10 AAGTAGCTCCCTGCTACTT 30

RESULT 3

R67982 417 bp mRNA EST 01-JUN-1995

LOCUS

Y104410.r1 Soares placenta Nb2HP Homo sapiens cDNA clone

DEFINITION

IMAGE:138234 5', mRNA sequence.

ACCESSION

R67982.1 GI:841499

VERSION

EST.

KEYWORDS

human.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE

1 (bases 1 to 417)

JOURNAL

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

COMMENT

, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maita, M., Parsons, J.,

Unpublished (1995)

Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston

The Mashu-Merck EST Project

, R., Williamson, A., Wohldmann, P. and Wilson, R.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estevason.wustl.edu

Insert Size: 774

High quality sequence stops: 306

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 774

Std Error: 0.00

Seq primer: M13RPI

Seg primer: M13RPI

High quality sequence stop: 306

Location/Qualifiers

1. 418

/organism="Homo sapiens"

/db_xref="GDB:540219"

/db_xref="taxon:9606"

/clone="IMAGE:134398"

/clone_lib="Soares placenta Nb2HP"

High quality sequence stop: 306.

Location/Qualifiers

1. 417

/organism="Homo sapiens"

/db_xref="GDB:544609"

/db_xref="taxon:9606"

/clone="IMAGE:138234"

/clone_lib="Soares placenta Nb2HP"

/sex="Female"

/dev_stage="Placenta obtained at birth (full term)"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: placenta; Vector: pT73D (Pharmacia) with a

modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

AACTGAGAGATTCGGCGCCGAGATTTTCTTTTCTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization. Library

constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT

104 a 101 c 89 g 123 t

Query Match 84.8%; Score 17.8; DB 188; Length 417;

Best Local Similarity 90.5%; Pred. No. 1.6e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aagtgactccctgactct 21

||||| |||||||

Db 10 AAGTAGCTCCCTGCTACTT 30

RESULT 4

R31714 418 bp mRNA EST 28-APR-1995

LOCUS

Yh63a12.r1 Soares placenta Nb2HP Homo sapiens cDNA clone

DEFINITION

IMAGE:134398 5', mRNA sequence.

ACCESSION

R31714.1 GI:787557

VERSION

EST.

KEYWORDS

human.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE

1 (bases 1 to 418)

JOURNAL

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

COMMENT

, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maita, M., Parsons, J.,

Unpublished (1995)

Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston

The Mashu-Merck EST Project

, R., Williamson, A., Wohldmann, P. and Wilson, R.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estevason.wustl.edu

Insert Size: 707

High quality sequence stops: 244

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 707

Std Error: 0.00

Seq primer: M13RPI

High quality sequence stop: 244.

Location/Qualifiers

1. 418

/organism="Homo sapiens"

/db_xref="GDB:540219"

/db_xref="taxon:9606"

/clone="IMAGE:134398"

/clone_lib="Soares placenta Nb2HP"

/sex="Female"
 /dev_stage="placenta obtained at birth (full term)"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Placenta; Vector: pRT3D (Pharmacia) with a modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACGAGAGAAATTCGCGCCGACAGAAATTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3D vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 90 a 100 c 95 g 126 t 7 others
 ORIGIN

Query Match 84.8%; Score 17.8; DB 187; Length 418;
 Best Local Similarity 90.5%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aagtgacccctgcctaccc 21
 |||||
 Db 10 AAGTACCTCCCTGCTACCTT 30

RESULT 5
 R63154 481 bp mRNA EST 26-MAY-1995
 LOCUS
 DEFINITION Y101a11.r1 Soares placenta NB2HP Homo sapiens cDNA clone
 IMAGE:137948 5', mRNA sequence.
 ACCESSION R63154
 VERSION R63154.1 GI:835033
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 481)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE The Washu-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson R.
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 Insert Size: 769
 High quality sequence stops: 411
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 Seg primer: M13RPI
 High quality sequence stop: 411.
 Location/Qualifiers

FEATURES

source

1. 481
 /organism="Homo sapiens"
 /db_xref="GDB:544291"
 /db_xref="taxon:9606"
 /clone="IMAGE:137948"
 /clone_1lb="Soares placenta NB2HP"
 /sex="Female"
 /dev_stage="placenta obtained at birth (full term)"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Placenta; Vector: pRT3D (Pharmacia) with a modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACGAGAGAAATTCGCGCCGACAGAAATTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3D vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 112 a 114 c 99 g 152 t 4 others
 ORIGIN

Query Match 84.8%; Score 17.8; DB 188; Length 481;
 Best Local Similarity 90.5%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aagtgacccctgcctaccc 21
 |||||
 Db 11 AAGTACCTCCCTGCTACCTT 31

RESULT 6
 A0882343 498 bp DNA GSS 09-NOV-1999
 LOCUS
 DEFINITION HS_5382.A1-B03-77C RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-9150 Col-5 Row-C, DNA sequence.
 ACCESSION A0882343
 VERSION A0882343.1 GI:6313810
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 498)
 AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@u.washington.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.bufileo.edu/ordering/bac.htm>) or from Research Genetics (info@resgen.com). BAC end Web Server: <http://www.hsc.washington.edu>
 Plate: 9150 row: C column: 5
 Seg primer: T7
 Class: BAC ends
 High quality sequence stop: 498.
 Location/Qualifiers

FEATURES

source

1. 498
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-9150 Col-5 Row-C"
 /clone_1lb="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBAC63.6; Site: 1: EcoRI; Site 2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBAC63.6 vector at EcoRI sites"
 BASE COUNT 155 a 111 c 118 g 111 t 3 others
 ORIGIN

Query Match 84.8%; Score 17.8; DB 235; Length 498;
 Best Local Similarity 90.5%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aagtgacctccctgacctt 21
 |||
 DB 74 AAGTGAACCTCCCTGCTT 94

RESULT 7
 A0185931/c 510 bp DNA GSS 01-NOV-1998
 LOCUS
 DEFINITION HS_3065_A1-D09_MF CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone plate-3065 COL-17 Row-G, DNA sequence.
 A0185931
 ACCESSION A0185931.1 GI:3585373
 VERSION
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 510)
 Mahatras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahatras GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence tagged Connector
 Plate: 3065 Row: G Column: 17
 Class: BAC ends
 High quality sequence stop: 510.
 Location/Qualifiers
 1..510
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="plate-3065 COL-17 Row-G"
 /clone_1lb="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBeloBAC11; BAC clones in
 E-Coli DH10B"
 BASE COUNT 157 a 98 c 100 g 153 t 2 others
 ORIGIN

Query Match 84.8%; Score 17.8; DB 225; Length 510;
 Best Local Similarity 90.5%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aagtgacctccctgacctt 21
 |||
 DB 159 AATGACACCTCGCTACTT 139

RESULT 8
 A0551067 608 bp DNA GSS 28-MAY-1999
 LOCUS
 DEFINITION RPCI-11-416M22.TV RPCI-11 Homo sapiens genomic clone RPCI-11-416M22
 / DNA sequence.
 A0551067
 ACCESSION A0551067.1 GI:4910244
 VERSION
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 608)
 Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter

TITLE Use of BAC End sequences from Library RPCI-11 for Sequence-Ready
 Map Building
 JOURNAL Unpublished (1997)
 COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbaetig@igf.org
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pletier de Jong
 (pletier@jng.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
 Research Genet cs (info@resgen.com). BAC end search page:
 http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..608
 /organism="Homo sapiens"
 /db_xref="GDB:7659669"
 /db_xref="taxon:9606"
 /clone="RPCI-11-416M22"
 /clone_1lb="RPCI-11"
 /sex="male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACes1.6; Site_1: EcoRI; Site_2: EcoRI;
 RPCI11 Human Male BAC library"
 BASE COUNT 179 a 152 c 134 g 143 t
 ORIGIN

Query Match 84.8%; Score 17.8; DB 230; Length 608;
 Best Local Similarity 90.5%; Pred. No. 1.7e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aagtgacctccctgacctt 21
 |||
 DB 376 AAGTGAACCTCCCTGCTT 396

RESULT 9
 AV553454 471 bp mRNA EST 07-SEP-2000
 LOCUS AV553454 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
 DEFINITION CDNA clone R263b05R 5', mRNA sequence.
 AV553454
 ACCESSION AV553454.1 GI:8724867
 VERSION
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 471)
 Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 DNA Res. 7, 175-180 (2000)
 CONTACT: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yata 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
 Location/Qualifiers
 1..471
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"

```

QY      5 gaactccctgtactct 21
          |||||
Db      350 GAACCTCCCTGTACTCT 330

```

RESULT	10
AM381670/c	
LOCUS	AM381670 193 bp mRNA EST 04-FEB-2000
DEFINITION	CVO-HR0310-061229-06g-g09 HT0310 Homo sapiens cDNA, mRNA sequence
ACCESSION	AM381670
VERSION	AM381670.1 GI:6886329
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	1 (bases 1 to 193)
AUTHORS	HCGP http://www.ludwig.org.br/CHESTES .
TITLE	The FAPESP/LICR Human Cancer Genome Project
JOURNAL	Unpublished (1999)
COMMENT	Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the PAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?lt=OV0&t2=OV0-HF0310-061259-065-g09&t3=1999-12-06&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 17
 High quality sequence stop: 193.
 Location/Qualifiers
 1..193

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0310"
/dev_stage="Adult"
/note="Organ: head,neck; Vector: puc18; Site:1; SmaI:
Site:2; SmaI: A mini-library was made by cloning products
derived from ORSTS PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

```
Query March      80.0%   Score 16.8 ; DB 115; Length 193;
Best Local Similarity 90.0%; Pred. NO. 4.4e-02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0.
```

RESULT	11
AA500281	
LOCUS	VA970708.1 304 bp mRNA EST 01-JUL-1997
DEFINITION	clone IMAGE:520199.5, mRNA sequence.
ACCESSION	AA500281
VERSION	AA500281.1
KEYWORDS	GI:2235248

ORGANISM

REFERENCE
AUTHORS

TITLE	JOURNAL	COMMENT
The Washu-HHMI Mouse EST Project	Unpublished (1996)	Contact: Marra M/Mouse EST Project

FEATURES

Email: mousteestwatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:332415
Seq primer: -26m13 rev2 ET from Amersham
High quality sequence stop: 225.
Location/analysts

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="920199"
/clone_id="Barstead mouse pooled organs MELRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: p773D-Pac (pharmacia) with:
a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (9
TGTTCAGATCTCAATGAGGAGCGCCGCCCTTTTTTTTTTTTTTTTTTTT
3'); double-stranded cDNA was ligated to Eco RI adaptors
(GTGGATCGTCGAC), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
Library constructed by Bob Barstead."

```

```

Query March      80.08; Score 16.8; DB 8; Length 304;
Best Local Similarity 90.08; Pzed No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy      1 aagtgaaactcctcgtctact 20
          |||||
Db      152 AAGTGAACTCCTGGTTCCT 171

```

RESULT	12
AA560193	
LOCUS	
DEFINITION	
ACCESSION	AA560193
VERSION	326 bp mRNA EST
KEYWORDS	18-AUG-1997
	V120901.r1 Strataene mouse Tcell j37311 Mus musculus cDNA clone
	IMAGE:72816 5', mRNA sequence.
	AA560193
	AA560193.1 GI.2331658
	EST.

SOURCE
house mouse.
ORGANISM
Mus musculus.
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 326)
AUTHORS
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Giesel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Stepcio, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:553544
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 294.
FEATURES
source
1. 326
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="972816"
/clone_id="Stratagene mouse Tcell 937111"
/tissue_type="Tcell"
/dev_stage="M30 CD4+ cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: blood; Vector: pBluescript SK-; Site: 1; EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer: 0.150 dt. M30 CD4+ cells. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGACGAC 3' -3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGTGTGT 3' "

BASE COUNT
71 a 89 c 89 g 76 t 1 others
ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 326;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aagtaactccctgact 20
|||||
Db 103 AAGTGAACCTCCTGTGCT 122

RESULT 13
AA620205 410 bp mRNA EST 09-OCT-1997
LOCUS v064a06.t1 Soares_mammary_gland_NBMWG Mus musculus cDNA clone
DEFINITION IMAGE:1054642 5', mRNA sequence.
ACCESSION AA620205
VERSION AA620205.1 GI:2524081
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 410)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Giesel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Stepcio, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT WashU-HMI Mouse EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:586218
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 400.
FEATURES
source
1. 410
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="1054642"
/clone_id="Soares_mammary_gland_NBMWG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site: 1; Not I; Site: 2; Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5']
TGTACCAATCTGAAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTT
T 3'; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773D vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
95 a 116 c 115 g 84 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 410;
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aagtaactccctgact 20
|||||
Db 8 AAGTGAACCTCCTGTGCT 27

RESULT 14
A1615361 416 bp mRNA EST 21-APR-1999
LOCUS v197f08.y1 Barstead mouse pooled organs MPRB4 Mus musculus cDNA
DEFINITION clone IMAGE:920193 5' similar to TR:075935 075935 DYNACTIN SUBUNIT. 1; mRNA sequence.
ACCESSION A1615361
VERSION A1615361.1 GI:4624528
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 416)
AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepcio, M., Theising, B., Allen, M., Bowles, M., Peterson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, J., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: July 25, 2001, 05:18:51 ; Search time 117.39 Seconds
(without alignments)
33.141 Million cell updates/sec

Title: US-09-142-095-2

Perfect score: 21
Sequence: 1 ccactgggatacaagttct 21

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/pdata/2/lna/5A_COMB.seq:*
2: /cgn2_6/pdata/2/lna/5B_COMB.seq:*
3: /cgn2_6/pdata/2/lna/6A_COMB.seq:*
4: /cgn2_6/pdata/2/lna/6B_COMB.seq:*
5: /cgn2_6/pdata/2/lna/PCTUS_COMB.seq:*
6: /cgn2_6/pdata/2/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
C 1	21	100.0	1190 5 PCT-US92-00282-18	Sequence 18, Appl
C 2	21	100.0	2336 5 PCT-US92-00282-1	Sequence 1, Appl
C 3	15.8	75.2	1073 4 US-08-960-780-37	Sequence 37, Appl
C 4	15.8	75.2	1073 4 US-08-073-898-37	Sequence 37, Appl
C 5	14.8	70.5	831 2 US-08-403-853-17	Sequence 17, Appl
C 6	14.8	70.5	832 1 US-08-473-496-1	Sequence 1, Appl
C 7	14.8	70.5	832 2 US-08-454-028-3	Sequence 3, Appl
C 8	14.8	70.5	832 5 PCT-US94-05388-3	Sequence 3, Appl
C 9	14.8	70.5	832 5 PCT-US95-09121-1	Sequence 1, Appl
C 10	14.8	70.5	832 5 PCT-US96-07496-3	Sequence 3, Appl
C 11	14.8	70.5	3225 1 US-08-698-551-13	Sequence 13, Appl
C 12	14.8	70.5	3225 2 US-08-602-228-13	Sequence 13, Appl
C 13	14.8	70.5	3225 2 US-08-533-901B-13	Sequence 13, Appl
C 14	14.8	70.5	3225 2 US-08-839-032A-13	Sequence 13, Appl
C 15	14.8	70.5	3225 2 US-08-839-032A-13	Sequence 13, Appl
C 16	14.8	70.5	3225 5 PCT-US95-12724-13	Sequence 13, Appl
C 17	14.6	69.5	87350 3 US-08-781-891-79	Sequence 19, Appl
C 18	14.2	67.6	20 4 US-08-669-378-16	Sequence 16, Appl
C 19	14.2	67.6	330 1 US-08-558-735-19	Sequence 19, Appl
C 20	14.2	67.6	330 4 US-08-906-480-19	Sequence 19, Appl
C 21	14.2	67.6	333 1 US-08-558-735-18	Sequence 18, Appl
C 22	14.2	67.6	333 4 US-08-906-480-18	Sequence 18, Appl
C 23	14.2	67.6	336 2 US-08-888-366-17	Sequence 17, Appl
C 24	14.2	67.6	343 1 US-08-320-559-36	Sequence 36, Appl
C 25	14.2	67.6	343 3 US-08-545-860D-36	Sequence 36, Appl
C 26	14.2	67.6	343 5 PCT-US94-04496-36	Sequence 36, Appl
C 27	14.2	67.6	345 1 US-08-558-735-3	Sequence 3, Appl

C 28	14.2	67.6	345 1	US-08-558-735-17	Sequence 17, Appl
C 29	14.2	67.6	345 4	US-08-906-480-3	Sequence 3, Appl
C 30	14.2	67.6	345 4	US-08-906-480-17	Sequence 17, Appl
C 31	14.2	67.6	335 1	US-08-558-735-16	Sequence 16, Appl
C 32	14.2	67.6	335 4	US-08-906-480-16	Sequence 16, Appl
C 33	14.2	67.6	428 3	US-08-589-939-4	Sequence 4, Appl
C 34	14.2	67.6	532 1	US-08-558-735-1	Sequence 1, Appl
C 35	14.2	67.6	532 4	US-08-906-480-1	Sequence 1, Appl
C 36	14.2	67.6	780 3	US-09-027-449-54	Sequence 54, Appl
C 37	14.2	67.6	780 3	US-09-027-449-58	Sequence 54, Appl
C 38	14.2	67.6	780 4	US-09-027-449-65	Sequence 54, Appl
C 39	14.2	67.6	780 4	US-08-804-44A-54	Sequence 54, Appl
C 40	14.2	67.6	780 4	US-08-804-44A-54	Sequence 54, Appl
C 41	14.2	67.6	780 4	US-09-026-985-54	Sequence 54, Appl
C 42	14.2	67.6	780 4	US-09-026-985-58	Sequence 54, Appl
C 43	14.2	67.6	780 4	US-09-026-985-65	Sequence 54, Appl
C 44	14.2	67.6	1422 4	US-08-387-117-8	Sequence 8, Appl
C 45	14.2	67.6	1905 4	US-08-387-117-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
PCT-US92-00282-18/C
Sequence 18, Application PC/TUS9200282
GENERAL INFORMATION:
APPLICANT: OMENS, IDA S.
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABRY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1190 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US92-00282-18

Query Match 100.0%; Score 21; DB 5; Length 1190;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 190 CCACCTGGATCAAGTTCCT 21

RESULT 2
PCT-US92-00282-1/c
Sequence 1, Application PC/TUS9200282
GENERAL INFORMATION:
APPLICANT: OENS, IDA S.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2336 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US92-00282-1

Query Match 100.0%, Score 21, DB 5, Length 2336;
Best Local Similarity 100.0%, Pred. No. 0.082;
Matches 21, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 1 ccaactggatcacagatct 21
DB 106 CCACGCGATCAACAGTACT 86

RESULT 3
US-08-960-780-37
Sequence 37, Application US/08960780
Patent No. 6204435
GENERAL INFORMATION:
APPLICANT: Feltelson, Jerald S.
APPLICANT: Schnepf, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Schmeltz, James
APPLICANT: Loewer, David
APPLICANT: Dullum, Charles Joseph
APPLICANT: Muller-Cohn, Judy
APPLICANT: Stamp, Lisa
TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/960,780
FILING DATE: 30-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA-708
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1073 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 196F3
US-08-960-780-37

Query Match 75.2%, Score 15.8, DB 4, Length 1073;
Best Local Similarity 89.5%, Pred. No. 29;
Matches 17, Conservative 0, Mismatches 2, Indels 0, Gaps 0;

QY 3 actggatcacagatct 21
DB 838 ACTGGCAAAAACAGTACT 856

RESULT 4
US-09-073-898-37
Sequence 37, Application US/09073898
Patent No. 6242669
GENERAL INFORMATION:
APPLICANT: Feltelson, Jerald S.
APPLICANT: Schnepf, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Schmeltz, James
APPLICANT: Loewer, David
APPLICANT: Dullum, Charles Joseph
APPLICANT: Muller-Cohn, Judy
APPLICANT: Stamp, Lisa
APPLICANT: Mortill, George
APPLICANT: Flisad-Lee, Stacey
TITLE OF INVENTION: No. 6242669el Pesticidal Toxins and Nucleotide
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073,898
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
PRIOR APPLICATION DATA: US 08/960,780
APPLICATION NUMBER: US 08/960,780
FILING DATE: 30-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sanders, Jay M.
REGISTRATION NUMBER: 39,355
REFERENCE/DOCKET NUMBER: MA-708C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1073 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 196F3
US-09-073-898-37

Query Match 75.2%; Score 15.8; DB 4; Length 1073;
Best Local Similarity 89.5%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 actggatcacagatctc 21
||||| |||||||
Db 838 ACTGGGAAAACAGTATCT 856

RESULT 5
US-08-403-853-17
Sequence 17, Application US/08403853
Patent No. 5844094
GENERAL INFORMATION:
APPLICANT: HUDSON, Peter J.
APPLICANT: LAH, Maria
APPLICANT: KORST, Alex A.
APPLICANT: IRVING, Robert A.
APPLICANT: ATWELL, John L.
APPLICANT: MALBY, Robyn L.
APPLICANT: POWER, Barbara E.
APPLICANT: COLMAN, Peter M.
TITLE OF INVENTION: TARGET BINDING POLYPEPTIDE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,853
FILING DATE: 30-MAY-1995
CLASSIFICATION: 433
PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/AU93/00491
FILING DATE: 24-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PL 4973
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16786/189/CHAC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 831 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1..819
NAME/KEY: CDS
LOCATION: 1..819
US-08-403-853-17

Query Match 70.5%; Score 14.8; DB 2; Length 831;
Best Local Similarity 88.9%; Pred. No. 88;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 actggatcacagatc 20
||||| |||||||
Db 575 ACTGGTACACAGATC 592

RESULT 6
US-08-473-496-1/c
Sequence 1, Application US/08473496
Patent No. 5700660
GENERAL INFORMATION:
APPLICANT: Jack L. Leonard
APPLICANT: Peter E. Newburger
TITLE OF INVENTION: POSITIONAL CONTROL OF SELENIUM INSERTION
TITLE OF INVENTION: IN POLYPEPTIDES FOR X-RAY CRYSTALLOGRAPHY
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,496
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/277,492
FILING DATE: 19-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/066,680
FILING DATE: 24-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 04020/078001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154

SEQUENCE CHARACTERISTICS:

LENGTH: 832 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-473-496-1

Query Match 70.5%; Score 14.8; DB 1; Length 832;
Best Local Similarity 88.9%; Pred. No. 88;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 actggatcaacagatc 20
DB 768 ACTGGATCAACAGACC 751

RESULT 7

US-08-454-028-3/c
Sequence 3, Application US/08454028
Patent No. 5849520

GENERAL INFORMATION:

APPLICANT: Jack L. Leonard
APPLICANT: Peter E. Newburger
TITLE OF INVENTION: POST-TRANSCRIPTIONAL GENE REGULATION BY
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,028
FILING DATE: May 30, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/066,680
FILING DATE: May 24, 1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 04020/075001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154

SEQUENCE CHARACTERISTICS:

LENGTH: 832 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-454-028-3

Query Match 70.5%; Score 14.8; DB 2; Length 832;
Best Local Similarity 88.9%; Pred. No. 88;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 actggatcaacagatc 20
DB 768 ACTGGATCAACAGACC 751

RESULT 8

PCT-US94-05388-3/c
Sequence 3, Application PC/TUS9405388
GENERAL INFORMATION:

APPLICANT: Jack L. Leonard
APPLICANT: Peter E. Newburger
TITLE OF INVENTION: POST-TRANSCRIPTIONAL GENE
REGULATION BY TRACE
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/05388
FILING DATE: 16 May 1994
CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/066,680
FILING DATE: May 24, 1993
ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 04020/026001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 832
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
PCT-US94-05388-3

Query Match 70.5%; Score 14.8; DB 5; Length 832;
Best Local Similarity 88.9%; Pred. No. 88;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 actggatcaacagatc 20
DB 768 ACTGGATCAACAGACC 751

RESULT 9

PCT-US95-09121-1/c
Sequence 1, Application PC/TUS9509121
GENERAL INFORMATION:

APPLICANT: University of Massachusetts
APPLICANT: Medical School
TITLE OF INVENTION: POSITIONAL CONTROL OF SILENIUM
TITLE OF INVENTION: INSERTION IN POLYPEPTIDES FOR
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09121
FILING DATE: 19 July 1995
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/277,492
FILING DATE: 19 July 1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/473,496
FILING DATE: 7 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 04020/078M01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
PCT-US95-09121-1

Query Match 70.5%; Score 14.8; DB 5; Length 832;
Best Local Similarity 88.9%; Pred. No. 88;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 actggatcaacagatc 20
DB 768 ACTGGATCAACAGACC 751

RESULT 10
PCT-US96-07496-3/C
Sequence 3, Application PC/TUS9607496
GENERAL INFORMATION:
APPLICANT: University of Massachusetts Medical Center
TITLE OF INVENTION: POST-TRANSCRIPTIONAL GENE REGULATION BY
TITLE OF INVENTION: SELENIUM
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/07496
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/454,028
FILING DATE: May 30, 1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/066,680
FILING DATE: May 24, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 04020/075M01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
PCT-US96-07496-3

Query Match 70.5%; Score 14.8; DB 5; Length 832;
Best Local Similarity 88.9%; Pred. No. 88;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 actggatcaacagatc 20
DB 768 ACTGGATCAACAGACC 751

RESULT 11
US-08-698-551-13/C
Sequence 13, Application US/08698551
Patent No. 5712361
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schiavella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/698,551
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15332D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 498-5851
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 3225 base pairs

NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,032A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232BDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 3225 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 3..2846
US-08-839-032A-13
Query Match 70.5%; Score 14.8; DB 2; Length 3225;
Best Local Similarity 88.9%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CY 4 ctggatcaacagttatct 21
||||| |||||||||
DB 504 CTGGGTCCAAACAGTATCT 487
RESULT 15
US-08-839-031A-13/C
Sequence 13, Application US/08839031A
Patent No. 5948638
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,031A

FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.,
REGISTRATION NUMBER: 41,323
REFERENCE/DOCKET NUMBER: G15232BDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 3225 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 3..2846
US-08-839-031A-13

Query Match 70.5%; Score 14.8; DB 2; Length 3225;
Best Local Similarity 88.9%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CY 4 ctggatcaacagttatct 21
||||| |||||||||
DB 504 CTGGGTCCAAACAGTATCT 487

Search completed: July 25, 2001, 05:18:52
Job time: 9215 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 04:54:49 ; Search time 2762.24 Seconds
(without alignments)
71.866 Million cell updates/sec

Title: US-09-142-095-2

Perfect score: 21

Sequence: 1 ccactggatcacacagatctc 21

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1022815 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*
259: em_gss_inv57:*
260: em_gss_inv58:*
261: em_gss_inv59:*
262: em_gss_inv60:*
263: em_gss_inv61:*
264: em_gss_inv62:*
265: em_gss_inv63:*
266: em_gss_inv64:*
267: em_gss_inv65:*
268: em_gss_inv66:*
269: em_gss_inv67:*
270: em_gss_inv68:*
271: em_gss_inv69:*
272: em_gss_inv70:*
273: em_gss_inv71:*
274: em_gss_inv72:*
275: em_gss_inv73:*
276: em_gss_inv74:*
277: em_gss_inv75:*
278: em_gss_inv76:*
279: em_gss_inv77:*
280: em_gss_inv78:*
281: em_gss_inv79:*
282: em_gss_inv80:*
283: em_gss_inv81:*
284: em_gss_inv82:*
285: em_gss_inv83:*
286: em_gss_inv84:*
287: em_gss_inv85:*
288: em_gss_inv86:*
289: em_gss_inv87:*
290: em_gss_inv88:*
291: em_gss_inv89:*
292: em_gss_inv90:*
293: em_gss_inv91:*
294: em_gss_inv92:*
295: em_gss_inv93:*
296: em_gss_inv94:*
297: em_gss_inv95:*
298: em_gss_inv96:*
299: em_gss_inv97:*
300: em_gss_inv98:*
301: em_gss_inv99:*
302: em_gss_inv100:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21	100.0	327	189	T71061	T71061 ycs0a04.r1
2	18.4	87.6	383	3	AA195421	AA195421 z33h09.s
3	18.4	87.6	1078	221	CNS04CBO	AL284229 Tetraodon
4	17.8	84.8	297	169	BF764030	BF764030 IL2-CS004
5	17.8	84.8	425	3	AA195418	AA195418 z33h09.s
6	17.8	84.8	425	3	AA195418	AA195418 z33h09.s
7	17.4	82.9	499	107	AA066229	AA066229 RPT-23-3
8	17.4	82.9	499	107	AA066229	AA066229 RPT-23-3
9	17.4	82.9	499	107	AA066229	AA066229 RPT-23-3
10	16.8	80.0	437	11	AA78238	AA78238 z19708.s
11	16.8	80.0	440	107	AA070697	AA070697 AA070697
12	16.8	80.0	450	107	AA094227	AA094227 AA094227
13	16.8	80.0	484	151	BP598564	BP598564 BP598564
14	16.8	80.0	520	155	BP557402	BP557402 BP557402
15	16.8	80.0	534	232	AA073479	AA073479 AA073479
16	16.8	80.0	535	145	BF193631	BF193631 BF193631
17	16.8	80.0	551	251	AA291013	AA291013 RPT-24-1
18	16.8	80.0	583	244	AA246376	AA246376 AA246376
19	16.8	80.0	628	6	AA392024	AA392024 AA392024
20	16.8	80.0	629	6	AA392019	AA392019 AA392019
21	16.8	80.0	698	107	AA094228	AA094228 AA094228
22	16.8	80.0	713	14	AA951996	AA951996 AA951996
23	16.8	80.0	881	150	BF578486	BF578486 BF578486
24	16.4	78.1	306	160	BP520416	BP520416 BP520416
25	16.4	78.1	346	120	AA764293	AA764293 AA764293
26	16.4	78.1	522	142	BE942800	BE942800 BE942800
27	16.4	78.1	568	142	BE942802	BE942802 BE942802
28	16.4	78.1	683	240	AZ260166	AZ260166 RPT-23-1
29	16.4	78.1	694	219	AG013064	AG013064 RPT-23-1
30	16.4	78.1	697	175	BC282828	BC282828 BC282828
31	16.4	78.1	1055	146	BP304360	BP304360 BP304360
32	16.4	78.1	1071	220	CNS02E0E	AA193127 Tetraodon
33	16.2	77.1	219	149	BF490633	BF490633 BF490633
34	16.2	77.1	306	131	BB336747	BB336747 BB336747
35	16.2	77.1	317	245	AZ504982	AZ504982 AA0445019
36	16.2	77.1	403	247	AZ647967	AZ647967 AA0445019
37	16.2	77.1	412	118	AA560028	AA560028 AA560028
38	16.2	77.1	416	172	BG019825	BG019825 BG019825
39	16.2	77.1	419	246	AZ604463	AZ604463 AA0445019
40	16.2	77.1	437	113	AA253561	AA253561 AA253561
41	16.2	77.1	439	23	AA655576	AA655576 AA655576
42	16.2	77.1	439	104	AA1970357	AA1970357 AA1970357
43	16.2	77.1	440	118	AA560489	AA560489 AA560489
44	16.2	77.1	443	171	BF932393	BF932393 BF932393
45	16.2	77.1	446	257	BP98792	BP98792 BP98792

ALIGNMENTS

RESULT 1
T71061/c
LOCUS
DEFINITION
T71061 ycs0a04.r1 Stratagene liver (#937224) Homo sapiens CDNA clone
IMAGE:84078.5' similar to gb:M57899 UDP-GLUCURONOSYLTRANSFERASE 1A
PRECUSOR, MICROSOFT (HUMAN); mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
EST
human.
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 327)
Haller, L., Lennon, G., Becker, M., Bonaldo, M.F., Chippelli, B.,
Chisore, S., Dietrich, N., Dubuque, T., Favellio, A., Gish, N., Hawkins,
M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Martis, E., Moore

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 603
High quality sequence stops: 290 Source: IMAGE Consortium, LLNL This
clone is available royalty-free through LLNL; contact the IMAGE
Consortium (image.llnl.gov) for further information.
Insert Length: 603 Std Error: 0.00
Seq primer: M13R1
High quality sequence stop: 290.
Location/Qualifiers
1..327
/organism="Homo sapiens"
/db_xref="Gene:501135"
/db_xref="taxon:9606"
/clone="IMAGE:84078"
/clone_lib="Stratagene liver (#937224)"
/sex="male"
/dev_stage="49 years old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: liver; Vector: pBluescript SK; Site: 1; EcoRI
site: 2; XhoI: Cloned unidirectionally. Primer: Oligo
dt. Hepatocytoma from normal male caucasian. Average insert
size: 1.1 kb; Uni-ZAP XR Vector; -3' adaptor sequence: 5'
CAATCGGACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTT 3' "

FEATURES

BASE COUNT
71 a 92 g 83 t 7 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 21; DB 189; Length 327;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULTS

RESULT 2
LOCUS
DEFINITION
T71061 ycs0a04.r1 Stratagene liver (#937224) Homo sapiens CDNA clone IMAGE:84078.5' similar to gb:M57899 UDP-GLUCURONOSYLTRANSFERASE 1A
PRECUSOR, MICROSOFT (HUMAN); mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
EST
human.
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 383)
Haller, L., Lennon, G., Becker, M., Bonaldo, M.F., Chippelli, B.,
Chisore, S., Dietrich, N., Dubuque, T., Favellio, A., Gish, N., Hawkins,
M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Martis, E., Moore

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@wason.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.lnl.gov) for further information. Insert Length: 1881 Std Error: 0.00 High quality sequence stop: 218.

FEATURES

source

Location/Qualifiers

1..383

/organism="Homo sapiens"

/db_xref="GDB:5427569"

/db_xref="taxon:9606"

/clone_image="665537"

/clone_lib="Scars_NhMPu_S1"

/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"

/note="Organ: mixed (see below); Vector: pT73D-Pac"

(Pharmacia) with a modified polylinker; site_1: Not I;

site_2: Eco RI; Equal amounts of plasmid DNA from three

normalized libraries (melanocyte 2NBH, pregnant uterus

2NBH, and fetal heart 2NBH19W) were mixed, and ss circles

were made in vitro. Following HAP purification, this DNA

was used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from pools of

5,000 clones made from the same 3 libraries. The pools

consisted of 1.M.A.G.E. clones 260232-265223,

340488-345479, and 484488-489479."

BASE COUNT 108 a 83 c 72 g 112 t 8 others

ORIGIN

Query Match

Best Local Similarity 87.6%; Score 18.4; DB 3; Length 383;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 1 ccactggatcaacagatct 21

Db 180 CCACGTGGATCAGACAGTTC 200

RESULT 3

LOCUS

CNS04CB0/c

DEFINITION

099M21 of library G from Tetraodon nigroviridis, genomic survey

sequence.

ACCESSION

AL284229

VERSION

AL284229.1

KEYWORDS

GSS; genome survey sequence.

SOURCE

Tetraodon nigroviridis.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorphi; Tetraodontiformes;

Tetraodontidae; Tetraodon.

REFERENCE

1 (bases 1 to 1078)

AUTHORS

Roest-Crollius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C.,

Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and

Weissenbach,J.

TITLE

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

Unpublished

2 (bases 1 to 1078)

REFERENCE

Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,

Bernot,A., Fitzames,C., Wincker,P., Brothier,P., Queller,F.,

Saurin,W. and Weissenbach,J.

AUTHORS

Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

Unpublished

3 (bases 1 to 1078)

REFERENCE

Genoscope.

JOURNAL

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

COMMENT

This sequence is a single read and was generated as part of a large

scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cng.fr/Tetraodon>.

FEATURES

source

Location/Qualifiers

1..1078

/organism="Tetraodon nigroviridis"

/db_xref="taxon:99883"

/clone_image="099M21"

/clone_lib="g"

/note="Genoscope sequence ID: COBG09AG11SP1-end ;

PUC-ori"

BASE COUNT 343 a 192 c 230 g 312 t 1 others

ORIGIN

Query Match

Best Local Similarity 87.6%; Score 18.4; DB 221; Length 1078;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 ccactggatcaacagatct 20

Db 945 CCACGTGGATCAGACAGTTC 926

RESULT 4

LOCUS

BF764030/c

DEFINITION

IL2-CS0048-301000-198-610 CS0048 Homo sapiens CDNA, mRNA sequence.

ACCESSION

BF764030

VERSION

BF764030.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 297)

AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Bittencourt,M.R.,

Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=IL2&ct=IL2-CS0048-301000-198-610&ts=2000-10-30&td=1>)

Seq primer: puc 18 forward

High quality sequence stop: 58.

FEATURES

source

Location/Qualifiers

1..297

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="CS0048"

/dev_stage="Adult"

/note="Organ: colon_est; Vector: puc18; Site.1: SmaI;

Site.2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 60 a 81 c 95 g 61 t
ORIGIN

Query Match 84.8% Score 17.8; DB 169; Length 297;
Best Local Similarity 90.5%; Pred. No. 96;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 128 CCCTGGGATCAGACATCT 108

RESULT 5
AA195418 425 bp mRNA EST 06-AUG-1997
LOCUS z136g10.s1 Soares.NhMpu.St Homo sapiens cDNA clone IMAGE:665538 3'
DEFINITION similar to contains element MERS repetitive element ; mRNA
SEQUENCE.

ACCESSION AA195418.1 GI:1785111
VERSION
KEYWORDS
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

1 (bases 1 to 425)
Hiller, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacey, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Stepien, M., Tan, F., Theising, B., White, Y., Wyllie,
T., Waterston, R., and Wilson, R.
WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8500, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu

TITLE

JOURNAL

COMMENT

This clone is available royalty-free through LMAN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1873 Std Error: 0.00
High quality sequence stop: 337.
Location/Qualifiers

FEATURES

SOURCE

1. 425
/organism="Homo sapiens"
/db_xref="GDB:5427570"
/db_xref="taxon:9606"
/clone="IMAGE:665538"
/clone_id="Soares.NhMpu.St"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"

/note="Organ: mixed (see below); Vector: pTT73D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NDHM, pregnant uterus
NBHPU, and fetal heart NBHMI99) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of 1 M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT 129 a 85 c 80 g 129 t 2 others
ORIGIN

Query Match 84.8% Score 17.8; DB 3; Length 425;
Best Local Similarity 90.5%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 1 CCCTGGGATCAGACATCT 21

Db 180 CCCTGGGATCAGACATCT 200

RESULT 6
A2066229/c 612 bp DNA GSS 30-MAR-2000
LOCUS RPCI-23-392B4.TV RPCI-23 Mus musculus genomic clone RPCI-23-392B4,
DEFINITION DNA sequence.
ACCESSION A2066229.1 GI:7357481
VERSION
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 612)
Zhao, S., Nierman, M., Feldblum, T., Malek, J., Shatsman, S., Akinet,
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.,
and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1998)
Contact: Shuying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@edj.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 392 Row: B Column: 4
Seq primer: 7
Class: BAC ends.
Location/Qualifiers

FEATURES

SOURCE

1. 612
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-392B4"
/clone_id="RPCI-23"
/sex="Female"
/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 176 a 121 c 116 g 199 t
ORIGIN

Query Match 84.8% Score 17.8; DB 237; Length 612;
Best Local Similarity 90.5%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 540 CCCCTGGATCAGACATCT 520

RESULT 7
A0030401/c 499 bp mRNA EST 19-OCT-1998
LOCUS A0030401 Rice cDNA from immature leaf including optical metastem
DEFINITION Oryza sativa cDNA clone EST1014.4; mRNA sequence.
ACCESSION A0030401

VERSION A0030401.1 GI:3763649
 EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Oryza sativa.
 Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriacridaceae; Oryzaceae; Oryza.
 1 (bases 1 to 499)
 Sasaki, T. and Yamamoto, K.
 Rice cDNA from immature leaf including apical meristem
 unpublished (1997)
 Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abrr.affrc.go.jp
 PROJECT = "RGP".

FEATURES
 source
 location/Qualifiers
 1..499
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"
 /clone="E51014.4Z"
 /clone_1lb="Rice cDNA from immature leaf including apical
 meristem"
 /dev_stage="Immature"
 /note="Organ: leaf; immature leaf including apical
 meristem (under long day condition)"

BASE COUNT 132 a 112 c 133 g 115 t 7 others
 ORIGIN

Query Match 82.3%; Score 17.4; DB 107; Length 499;
 Best Local Similarity 90.0%; Pred. No. 1.7e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ccaactggatcaacgctac 20
 |||||
 DB 78 CCACGTCGATCAACAGTTTC 59

RESULT 8
 BE700745 624 bp mRNA EST 22-DEC-2000
 LOCUS BE700745
 DEFINITION 602128542F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4285378 5',
 mRNA sequence.
 ACCESSION BE700745
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 624)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@pds-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/MLNL at:
 http://image.llnl.gov
 Plate: LCM1120 row: c column: 11
 High quality sequence stop: 623.
 Location/Qualifiers

FEATURES

Source

1..624
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4285378"
 /clone_1lb="NIH_MGC_56"
 /tissue_type="primitive neuroectoderm"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: Brain; Vector: pDNR-LIB (Clontech); site:1:
 SfiI (ggccgctggcgc); site:2: SfiI (ggccattggcgc);
 Double stranded cDNA was prepared from cell line RNA. 5'
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGGCATTATGGCC-3' and 3' adaptor
 sequence: 5'-ATCTAGAGCGCGAGCGCGCAGCATG-dt(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."

BASE COUNT 125 a 175 c 182 g 142 t
 ORIGIN

Query Match 81.0%; Score 17; DB 168; Length 624;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccaactggatcaacgctac 17
 |||||
 DB 351 CCACGTCGATCAACAGT 367

RESULT 9
 AA778238 437 bp mRNA EST 05-FEB-1998
 LOCUS AA778238
 DEFINITION z197908.91 Soares.fetal.liver.spleen.INFUS.S1 Homo sapiens cDNA
 clone IMAGE:448766 3', mRNA sequence.
 ACCESSION AA778238
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 437)
 Hillier, L., Allen, M., Bowles, L., Dubnue, T., Geisel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
 White, Y., Wylie, T., Waters, R., and Wilson, R.
 White, Y., Wylie, T., Waters, R., and Wilson, R.
 WashU-NCI human EST Project
 Unpublished (1997)
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through MLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 399.
 Location/Qualifiers

FEATURES
 source

1..437
 /organism="Homo sapiens"
 /db_xref="GDB:1353023"
 /db_xref="taxon:9606"
 /clone="IMAGE:448766"
 /clone_1lb="Soares.fetal.liver.spleen.INFUS.S1"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and spleen; Vector: p773D (Pharmacia)
 with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
 This is a subtracted version of the original Soares fetal

liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACTGGAAGAATAATTAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p773 vector. Library constructed through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 116 a 104 c 99 g 118 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 11; Length 437;
Best Local Similarity 90.0%; Pred. NO. 3.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatct 21
|||||
Db 315 CACTGGGATCAACATATTT 296

RESULT 10
AU070697/c 440 bp mRNA EST 10-JUN-1999
LOCUS
DEFINITION AU070697 Rice cDNA from young root Oryza sativa cDNA clone
R10115_1A, mRNA sequence.
ACCESSION AU070697
VERSION AU070697
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
AUTHORS 1 (bases 1 to 440)
TITLE Yamamoto, K. and Sasaki, T.
JOURNAL Rice cDNA from young root
COMMENT Unpublished (1999)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT "RGP".

FEATURES
Source Location/Qualifiers
1..440
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="R10115_1A"
/clone_lib="Rice cDNA from young root"
/tissue_type="young root"

BASE COUNT 126 a 86 c 94 g 132 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 107; Length 440;
Best Local Similarity 90.0%; Pred. NO. 3.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatct 21
|||||
Db 29 CACTGGGATCAGCATATCT 10

RESULT 11
AU094227/c 450 bp mRNA EST 30-JUN-2000
LOCUS
DEFINITION AU094227 Rice panicle at flowering stage Oryza sativa cDNA clone

ACCESSION AU094227
VERSION AU094227.1 GI:8856909
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
AUTHORS 1 (bases 1 to 450)
TITLE Rice cDNA from panicle at flowering stage (2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT "RGP".

FEATURES
source Location/Qualifiers
1..450
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E3685"
/clone_lib="Rice panicle at flowering stage"
/dev_stage="flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 117 a 91 c 121 g 120 t 1 others
ORIGIN

Query Match 80.0%; Score 16.8; DB 107; Length 450;
Best Local Similarity 90.0%; Pred. NO. 3.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatct 21
|||||
Db 254 CACTGGGATCAGCATATCT 235

RESULT 12
BF598564/c 484 bp mRNA EST 12-DEC-2000
LOCUS
DEFINITION BF598564 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
sv19a09.y1 Gm-cl057 similar to SW:PLAS_LYC5S P17340 PLASTOCYANIN
Gm-cl057-17 5' similar to SW:PLAS_LYC5S P17340 PLASTOCYANIN
PRECUSOR. [1] ; mRNA sequence.

ACCESSION BF598564
VERSION BF598564.1 GI:11690888
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids 1; Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 484)

AUTHORS Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Corvelli, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 for further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com

Trace considered overall poor quality

High quality sequence stop: 1.

FEATURES

source

Location/Qualifiers

1. 484

/organism="Glycine max"

/db.xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: GM-cl057-17"

/clone.lib="GM-cl057"

/tissue_type="degenerating cotyledons, 2 week old seedling"

/lab_host="DH10B"

/note="Vector: Bluescript II SK⁺ Site:1: EcoRI; Site:2: XhoI; The cDNA library was constructed from mRNA isolated from degenerating cotyledons of 2 week old seedlings from p1468916. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

BASE COUNT 117 a 133 c 121 g 113 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 151; Length 484;
Best Local Similarity 90.0%; Pred. No. 3.4e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ccactgggatacagatc 20

|||||

DB 435 CCATGGCATCAACAGTATC 416

RESULT 13

BS557402

LOCUS

DEFINITION

VERSION

KEYWORDS

SOURCE

ORGANISM

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

BS557402.1

GI:13586400

EST.

sequence.

BS557402

BS557402

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cactggatcacagatct 21
 ||||| ||||| ||
 Db 210 CACTGGAGCAGACTACT 191

RESULT 15

BF193631 535 bp mRNA EST 02-NOV-2000
 LOCUS 245281 MARC 2Pig sus scrofa cDNA 5', mRNA sequence.
 DEFINITION
 ACCESSION BF193631
 VERSION BF193631.1 GI:11077000
 KEYWORDS EST.
 SOURCE
 ORGANISM

Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 1 (bases 1 to 535)
 Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
 Stone, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.W.
 and Keele, J.W.
 Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
 Unpublished (2000)
 Contact: Smith TPL

JOURNAL COMMENT

USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smithemall.marc.usda.gov
 Single pass sequencing. Bases called and alt-trimmed with phred
 v0.980904.e. Vector identified by cross-match with the -minscore 18
 and -mismatch 12 options.
 PCR Primers

FORWARD: AGGAACAGCTATGACAC
 BACKWARD: GTTTCAGCTACAGAC
 Plate: 74 row: K column: 12
 Seq primer: ATTAGTGACACTATAG.
 Location/Qualifiers
 1. 535

FEATURES
 source
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 2Pig"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI;
 Library made from pooled tissue from testis, ovary,
 endometrium, hypothalamus, pituitary, and placenta."
 BASE COUNT 163 a 136 c 118 g 118 t
 ORIGIN

Query Match 80.0%; Score 16.8; DB 145; Length 535;
 Best local similarity 90.0%; Pred. NO. 3.5e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccactggatcacagatc 20
 ||||| ||||| ||
 Db 514 CCACTGGATCAGACAGAC 533

Search completed: July 25, 2001, 04:54:51
 Job time: 10424 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK: nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:16:58 ; Search time 1290.33 Seconds

(without alignments)
239.749 Million cell updates/sec

Title: US-09-142-095-3

Perfect score: 20

Sequence: 1 gtcacgtgacacgtcaaac 20

Scoring table: IDENTITY_NTC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 segs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_pat1:*

2: gb_da2:*

3: gb_da3:*

4: gb_in1:*

5: gb_in2:*

6: gb_in3:*

7: gb_in4:*

8: gb_in5:*

9: gb_in6:*

10: gb_in7:*

11: gb_in8:*

12: gb_in9:*

13: gb_in10:*

14: gb_in11:*

15: gb_in12:*

16: gb_in13:*

17: gb_in14:*

18: gb_in15:*

19: gb_in16:*

20: gb_in17:*

21: gb_in18:*

22: gb_in19:*

23: gb_in20:*

24: gb_in21:*

25: gb_in22:*

26: gb_in23:*

27: gb_in24:*

28: gb_in25:*

29: gb_in26:*

30: gb_in27:*

31: gb_in28:*

32: gb_in29:*

33: gb_in30:*

34: gb_in31:*

35: gb_in32:*

36: gb_in33:*

37: gb_in34:*

38: gb_in35:*

39: gb_in36:*

40: gb_in37:*

41: gb_in38:*

42: gb_in39:*

43: gb_in40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	20	100.0	20 9 A65502	A65502 Sequence 3
2	20	100.0	531 89 AF352795	AF352795 Homo sapi
3	20	100.0	541 89 AF180372	AF180372 Homo sapi
4	20	100.0	620 9 A65504	A65504 Sequence 5
5	20	100.0	3341 91 D87674	D87674 Homo sapien
6	20	100.0	68770 86 AC006985	AC006985 Homo sapi
7	20	100.0	116619 70 AC026497	AC026497 Homo sapi
8	20	100.0	139872 89 AF297093	AF297093 Homo sapi

9 18.4 92.0 200 89 AF135466
 10 18.4 92.0 208 89 AF135467
 11 18.4 92.0 234 89 AF135462
 12 18.4 92.0 234 89 AF135463
 13 17.4 87.0 4201 94 AF135463
 14 17.4 87.0 99123 12 AF069276
 15 16.8 84.0 207 89 AF135470
 16 16.8 84.0 208 89 AF135464
 17 16.8 84.0 208 89 AF135468
 18 16.8 84.0 208 89 AF135469
 19 16.8 84.0 208 89 AF135471
 20 16.8 84.0 208 89 AF135471
 21 16.8 84.0 208 89 AF135471
 22 16.8 84.0 208 89 AF135471
 23 16.8 84.0 208 89 AF135471
 24 16.8 84.0 208 89 AF135471
 25 16.8 84.0 208 89 AF135471
 26 16.8 84.0 208 89 AF135471
 27 16.8 84.0 208 89 AF135471
 28 16.8 84.0 208 89 AF135471
 29 16.8 84.0 208 89 AF135471
 30 16.8 84.0 208 89 AF135471
 31 16.8 84.0 208 89 AF135471
 32 16.8 84.0 208 89 AF135471
 33 16.8 84.0 208 89 AF135471
 34 16.8 84.0 208 89 AF135471
 35 16.8 84.0 208 89 AF135471
 36 16.8 84.0 208 89 AF135471
 37 16.8 84.0 208 89 AF135471
 38 16.8 84.0 208 89 AF135471
 39 16.8 84.0 208 89 AF135471
 40 16.8 84.0 208 89 AF135471
 41 16.8 84.0 208 89 AF135471
 42 16.8 84.0 208 89 AF135471
 43 16.8 84.0 208 89 AF135471
 44 16.8 84.0 208 89 AF135471
 45 15.8 79.0 188 7 AB010592

ALIGNMENTS

RESULT 1
 A65502 20 bp DNA
 LOCUS A65502
 DEFINITION Sequence 3 from Patent WO9732042.
 ACCESSION A65502
 VERSION A65502.1 GI:4531237
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 20)
 AUTHORS Burchell, B.
 TITLE DRUG TRIAL ASSAY SYSTEM
 JOURNAL Patent: WO 9732042-A 3 04-SEP-1997;
 UNITY DUNDEE (GB)
 COMMENT Other publication AU 2224197 19970916.
 FEATURES
 source
 1. 20
 /organism="unidentified"
 /db_xref="taxon:32644"

BASE COUNT 7 a 6 c 4 g 3 t
 ORIGIN

Query Match 100.0%; Score 20; DB 9; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.65;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtcaagtgacacgtaaac 20
 Db 1 gtcaagtgacacgtaaac 20

RESULT 2
 LOCUS AF352795
 DEFINITION AF352795 531 bp DNA
 ACCESSION AF352795
 VERSION AF352795.1 GI:13569708
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 531)
 AUTHORS McKie, K., Kutlar, F., Glendenning, M. and Kutlar, A.
 TITLE 7 (TR) repeat polymorphism of the TATA box of human bilirubin
 UDP-glucuronosyltransferase 1-1 (UGT1A1) gene in a patient with
 sickle cell anemia + high bilirubinemia
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 531)
 AUTHORS McKie, K., Kutlar, F., Glendenning, M. and Kutlar, A.
 TITLE Direct Submission
 JOURNAL Submitted (23-FEB-2001) Medicine/Hemoglobin DNA Laboratory, Medical
 College of Georgia, 15th St., AC-1000, Augusta, GA 30912, USA
 FEATURES
 source
 1. 531
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2q37"
 /cell_type="white blood cell"
 /tissue_type="blood"
 /note="UGT1A1"
 /note="UGT1A1"
 /allele="UGT1A1*1"
 <1. 313
 /gene="UGT1A1"
 279. 295
 /gene="UGT1A1"
 279. 295
 /note="repeat polymorphism compared to UGT1A1 sequence
 presented in GenBank Accession Number AF180372; contains 7
 ta repeats frequently detected in patients with high
 bilirubinemia"
 /rpt_type="candem
 /rpt_type="candem
 314. 531
 /gene="UGT1A1"
 /product="bilirubin UDP-glucuronosyltransferase 1-1"
 334. 531
 /gene="UGT1A1"
 /codon_start=1
 /product="bilirubin UDP-glucuronosyltransferase 1-1"
 /protein_id="AAK31204.1"
 /db_xref="GI:13569709"
 /translation="MAVESQGRPLVLLLCVGPVSHAKILLIPVDSHMLSKL
 GAIQDLQQRHETVLAQDASL"

BASE COUNT 120 a 121 c 137 g 153 t
 ORIGIN

Query Match 100.0%; Score 20; DB 89; Length 531;
 Best Local Similarity 100.0%; Pred. No. 0.36;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtcaagtgacacgtaaac 20
 Db 228 gtcaagtgacacgtaaac 247

RESULT 3

```

AF180372      541 bp      DNA      PRI      05-OCT-1999
LOCUS      AF180372      Homo sapiens bilirubin UDP-glucuronosyltransferase 1-1 (UGT1) gene,
DEFINITION      UGT1*1 allele, partial cds.
ACCESSION      AF180372
VERSION      AF180372.1 GI:6010649
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 541)
AUTHORS      Kutlar,F., Stromek,E., Leitner,C., Nechtman,J. and Kutlar,A.
TITLE      Deletion of the TATA box polymorphism of the human bilirubin
JOURNAL      UDP-glucuronosyltransferase 1-1 gene (UGT1*1) in a patient with
              sickle cell anemia
              Unpublished
REFERENCE      2 (bases 1 to 541)
AUTHORS      Kutlar,F., Stromek,E., Leitner,C., Nechtman,J. and Kutlar,A.
TITLE      Direct Submission
JOURNAL      Submitted (24-AUG-1999) Medicine, Hematology/Oncology-Sickle Cell
              Center, Medical College of Georgia, 15th Street, AC-1000, Augusta,
              GA 30912, USA
FEATURES
  source      1..541
              location/Qualifiers
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="2"
              /map="2q37"
              /cell_type="white blood cells"
              /tissue_type="whole blood"
              1..>541
  gene        /gene="UGT1"
              /note="GNT1; UGT1A1"
              /allele="UGT1*1"
              1..313
  promoter    /gene="UGT1"
              226
  variation    /gene="UGT1"
              279..288
              /replace="t"
  repeat_region /note="Polymorphic region"
              /rpl_type=tandem
              279..291
  TATA_signal /gene="UGT1"
              /note="Ritter,J.K., et al., 1992, J. Biol. Chem.,
              267:3257-3261"
              314..>541
  mRNA        /gene="UGT1"
              /product="bilirubin UDP-glucuronosyltransferase 1-1"
              330..>541
  CDS         /gene="UGT1"
              /note="UDP glucosyltransferase 1"
              /codon_start=1
              /product="bilirubin UDP-glucuronosyltransferase 1-1"
              /protein_id="AA01205.1"
              /db_xref="GI:6010650"
              /translation="MAVESQSGRPVZGLLCLVLPVSHACKILLIPVDGSHLWML
              GAIDLOORHEIVLAPASLYIRDG"
  BASE COUNT      123 a      124 c      141 g      153 t
  ORIGIN
Query Match      100.0%; Score 20; DB 89; Length 541;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gtccactgacacgacacac 20
DB      228 gtccactgacacgacacac 247

RESULT      4
LOCUS      A65504
DEFINITION      A65504      620 bp      DNA      PAT      29-MAR-1999
              Sequence 5 from Patent WO9732042.
ACCESSION      A65504
VERSION      A65504.1 GI:4531239
KEYWORDS
SOURCE      unidentified.
ORGANISM      unidentified.
REFERENCE      1 (bases 1 to 620)
AUTHORS      Burchell,B.
TITLE      DRUG TRIAL ASSAY SYSTEM
JOURNAL      Patent: WO 9732042-A 5 04-SEP-1997;
              UNIV DUNDEE (GB)
              Other publication AU 2224197 19970916.
COMMENT      location/Qualifiers
FEATURES
  source      1..620
              /organism="unidentified"
              /db_xref="taxon:32644"
  BASE COUNT      157 a      127 c      151 g      185 t
  ORIGIN
Query Match      100.0%; Score 20; DB 9; Length 620;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gtccactgacacgacacac 20
DB      508 gtccactgacacgacacac 527

RESULT      5
LOCUS      D87674
DEFINITION      D87674      3341 bp      DNA      PRI      14-APR-2000
              Homo sapiens gene for bilirubin UDP-glucuronosyltransferase 1,
              promoter region and partial cds.
ACCESSION      D87674
VERSION      D87674.1 GI:3059176
KEYWORDS      bilirubin UDP-glucuronosyltransferase 1.
SOURCE      Homo sapiens DNA.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (sites)
AUTHORS      Ueyama,H., Koike,O., Soeda,Y., Sato,H., Satoh,Y., Okubo,I. and
              Doi,Y.
TITLE      Analysis of the promoter of human bilirubin
              UDP-glucuronosyltransferase gene (UGT1*1) in relevance to Gilbert's
              syndrome
              Hepatol. Res. 9, 152-163 (1997)
JOURNAL      2 (bases 1 to 3341)
REFERENCE      Ueyama,H.
AUTHORS      Direct Submission
JOURNAL      Submitted (04-SEP-1996) to the DDBJ/EMBL/GenBank databases. Hisao
              Ueyama, Shiga University of Medical Science, Department of Medical
              Biochemistry, Shiga, Otsu, Shiga 520-21, Japan (Tel:077-548-2162,
              Fax:077-548-2164)
COMMENT      Sequence updated (08-Jan-1997) by: Hisao Ueyama.
FEATURES
  source      1..3341
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="2q37"
              2538..2544
              /note="XRE"
              2606..2610
              /note="XRE"
              3088..3097
              /note="E-box"
              3101..3113
              /note="HNF-1 site"
  misc_feature      2606..2610
  misc_feature      3088..3097
  misc_feature      3101..3113
  misc_feature      /note="HNF-1 site"

```

CAAT_signal 3125..3129
TATA_signal 3149..3153
exon 3177..3341
gene 3192..3341
CDS 3192..3341
/number=1
/gene="UGT1*1"
/gene="UGT1*1"
/codon_start=1
/product="bilirubin UDP-glucuronosyltransferase 1"
/protein_id="BA25600.1"
/db_xref="GI:3059177"
/translation="WAYSSGGRRPLVGLLCLVGPVVSAGKILLIPVDSHWLSML
GAIQD"

BASE COUNT 893 a 695 c 803 g 950 t
ORIGIN

Query Match 100.0%; Score 20; DB 91; Length 3341;
Best local similarity 100.0%; Pred. No. 0.35;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtcacgtgacacagtcacac 20
|||||
Db 3088 gtcacgtgacacagtcacac 3107

RESULT 6
AC006985/c 68770 bp DNA PRI 21-DEC-1999
LOCUS AC006985 Homo sapiens BAC clone RP11-154L24 from 2, complete sequence.
DEFINITION AC006985
VERSION AC006985.2 GI:5732165
KEYWORDS
SOURCE
human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
MEDLINE Genome Res. 8 (11), 1097-1108 (1998)
REFERENCE 99063792

REFERENCE

AUTHORS Gattung, S., Stoneking, T. and Davidson, T.
TITLE The sequence of Homo sapiens BAC clone RP11-154L24
JOURNAL Unpublished
3 (bases 1 to 68770)

REFERENCE

AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE

AUTHORS 4 (bases 1 to 68770)
TITLE Direct Submission
JOURNAL Submitted (22-OCT-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE

AUTHORS 5 (bases 1 to 68770)
TITLE Direct Submission
JOURNAL Submitted (13-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE

AUTHORS 6 (bases 1 to 68770)
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE

AUTHORS 6 (bases 1 to 68770)
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE

AUTHORS 6 (bases 1 to 68770)
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE

AUTHORS 6 (bases 1 to 68770)
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE

AUTHORS 6 (bases 1 to 68770)
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE

AUTHORS 6 (bases 1 to 68770)
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE

Web site: <http://genome.wustl.edu/gsc>
Contact: saplens@wustl.edu
----- Summary Statistics
Center project name: H.NH0154124

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RP11-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Moon, E.Y., Zhao, B., Firengen, E.,
Tateno, M., Catalanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(<http://bacpac.med.buffalo.edu>)
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-332L11, 200 bp overlap.
Actual start of this clone is at base position 86134 of
RP11-332L11; actual end is at base position 68770 of RP11-154L24.

The clone RP11-154L24 contains a tandem repeat from base positions
86234 to 39039, this region contains some low quality data. The
assembly is consistent with the restriction digest information.

FEATURES

1. 68770
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-154L24"
/clone_lib="RP11-11"
3. 192
repeat_region
/rpt_family="L1"
221..528
repeat_region
/rpt_family="Alu"
529..1634
repeat_region
/rpt_family="L1"
1635..1878
repeat_region
/rpt_family="MIR"
2079..2271
repeat_region
/rpt_family="MER1-type"
2276..2336
repeat_region
/rpt_family="MER1-type"
4683..4806
repeat_region
/rpt_family="MIR"
5131..5389
repeat_region
/rpt_family="MALR"
8738..8871
repeat_region
/rpt_family="MIR"
9092..9371
repeat_region
/rpt_family="Achoho"
9967..10582
repeat_region
/rpt_family="L1"

```

repeat_region 10576..11200 /rpt_family="L1"
repeat_region 11196..11469 /rpt_family="L1"
repeat_region 11540..11654 /rpt_family="L1"
repeat_region 12108..12523 /rpt_family="L1"
repeat_region 13113..13149 /rpt_family="MALR"
repeat_region 13150..13171 /rpt_family="MERL_type"
repeat_region 13172..13286 /rpt_family="(CAG)n"
repeat_region 13338..13497 /rpt_family="MERL_type"
repeat_region 14727..14848 /rpt_family="MERL_type"
repeat_region 15890..16244 /rpt_family="MIR"
repeat_region 16380..16723 /rpt_family="MALR"
repeat_region 17483..17560 /rpt_family="L1"
repeat_region 17655..17747 /rpt_family="L2"
repeat_region 18220..18337 /rpt_family="L2"
repeat_region 18770..18816 /rpt_family="MERL_type"
repeat_region 19103..19121 /rpt_family="(CAG)n"
repeat_region 19122..19636 /rpt_family="AT-rich"
repeat_region 19637..19662 /rpt_family="L1"
repeat_region 19825..20014 /rpt_family="AT-rich"
repeat_region 20034..20098 /rpt_family="L1"
repeat_region 20254..21067 /rpt_family="MIR"
repeat_region 21141..21182 /rpt_family="L1"
repeat_region 24649..25018 /rpt_family="A-rich"
repeat_region 25258..25447 /rpt_family="Retroviral"
repeat_region 25777..25824 /rpt_family="MIR"
repeat_region 28020..28178 /rpt_family="(CAG)n"
repeat_region 28178..28222 /rpt_family="MIR"
repeat_region 28503..28558 /rpt_family="L2"
repeat_region 28558..28741 /rpt_family="L1"
repeat_region 28970..29064 /rpt_family="L1"
repeat_region 29276..29434 /rpt_family="L1"
repeat_region 29554..29712 /rpt_family="MERL_type"
repeat_region 30867..31015 /rpt_family="MERL_type"
repeat_region 31268..31682 /rpt_family="MERL_type"
repeat_region 33195..33390 /rpt_family="MALR"
repeat_region 33607..33760 /rpt_family="MIR"
repeat_region 33815..34127 /rpt_family="L2"

```

```

repeat_region /rpt_family="Alu"
repeat_region /rpt_family="GA-rich"
repeat_region /rpt_family="L2"
repeat_region /rpt_family="L2"
repeat_region /rpt_family="L2"
repeat_region /rpt_family="GA-rich"

```

Query Match 100.0% Score 20; DB 86; Length 68770;
 Best Local Similarity 100.0%; Pred. No. 0.5;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 gtcacgtgacacgtcacaac 20
Db 66259 gtcacgtgacacgtcacaac 66240

```

```

RESULT 7
AC026497 176619 bp DNA HTG 22-MAR-2000
LOCUS Homo sapiens chromosome 11 clone RP11-689A10 map 11, *** SEQUENCING
DEFINITION IN PROGRESS *** 43 unordered pieces.
ACCESSION AC026497.1 GI:7283186
VERSION HTG: HTGS_PHASE1.
KEYWORDS human.
SOURCE ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

```

```

REFERENCE 1 (bases 1 to 176619)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Bogdanovskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choe, P., Colangelo, M., Collins, S.,
Collins, A., Cooke, P., Dearlano, K., Dewar, J., Diaz, J.S.,
Dodgson, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gege, D.,
Galeano, J., Gardina, S., Glend, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heald, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehotzky, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McKean, P., McGuire, A., McKernan, K., McPherson, R.,
Meldrum, J., Menes, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, T., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, R., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, T.,
Tasfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggiano, D.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

TITLE
 JOURNAL
 COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project information
 Center project name: L6277
 Center clone name: 689_A10

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 43 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 * 1334 1433: contig of 1333 bp in length
 * 1434 2787: contig of 1354 bp in length
 * 2788 3998: contig of 1111 bp in length
 * 3999 4098: gap of 100 bp
 * 4099 5777: contig of 1679 bp in length
 * 5778 7685: gap of 100 bp
 * 7686 7785: contig of 1808 bp in length
 * 7786 8176: gap of 100 bp
 * 8177 8276: contig of 391 bp in length
 * 8277 9926: contig of 1650 bp in length
 * 9927 10026: gap of 100 bp
 * 10027 11499: contig of 1473 bp in length
 * 11500 11599: gap of 100 bp
 * 11600 13616: contig of 2017 bp in length
 * 13617 13716: gap of 100 bp
 * 13717 16073: contig of 2357 bp in length
 * 16074 16173: gap of 100 bp
 * 16174 18510: contig of 2337 bp in length
 * 18511 18610: gap of 100 bp
 * 18611 20495: contig of 1885 bp in length
 * 20496 20595: gap of 100 bp
 * 20596 23548: contig of 2953 bp in length
 * 23549 23648: gap of 100 bp
 * 23649 25910: contig of 2262 bp in length
 * 25911 26010: gap of 100 bp
 * 26011 28060: contig of 2050 bp in length
 * 28061 28160: gap of 100 bp
 * 28161 30134: contig of 1974 bp in length
 * 30135 30234: gap of 100 bp
 * 30235 32646: contig of 2412 bp in length
 * 32647 32746: gap of 100 bp
 * 32747 35096: contig of 2350 bp in length
 * 35097 35196: gap of 100 bp
 * 35197 37696: contig of 2500 bp in length
 * 37697 37796: gap of 100 bp
 * 37797 40539: contig of 2743 bp in length
 * 40540 40639: gap of 100 bp
 * 40640 45518: contig of 4879 bp in length
 * 45519 45618: gap of 100 bp
 * 45619 48728: contig of 3110 bp in length
 * 48729 48828: gap of 100 bp
 * 48829 52099: contig of 3271 bp in length
 * 52100 52199: gap of 100 bp
 * 52200 56747: contig of 4548 bp in length
 * 56748 56847: gap of 100 bp
 * 56848 60467: contig of 3620 bp in length
 * 60468 60567: gap of 100 bp
 * 60568 64947: contig of 4380 bp in length
 * 64948 65047: gap of 100 bp
 * 65048 69431: contig of 4384 bp in length
 * 69432 69531: gap of 100 bp
 * 69532 74167: contig of 4636 bp in length
 * 74168 74267: gap of 100 bp
 * 74268 78188: contig of 3921 bp in length
 * 78189 78288: gap of 100 bp
 * 78289 81664: contig of 3376 bp in length
 * 81665 81764: gap of 100 bp
 * 81765 85169: contig of 3405 bp in length
 * 85170 85269: gap of 100 bp
 * 85270 90445: contig of 5176 bp in length
 * 90446 90545: gap of 100 bp
 * 90546 95008: contig of 4463 bp in length

* 95009 95108: gap of 100 bp
 * 95109 100587: contig of 5479 bp in length
 * 100588 100687: gap of 100 bp
 * 100688 105884: contig of 5197 bp in length
 * 105885 105984: gap of 100 bp
 * 105985 111486: contig of 5502 bp in length
 * 111487 111586: gap of 100 bp
 * 111587 117881: contig of 6195 bp in length
 * 117882 117881: gap of 100 bp
 * 117882 125343: contig of 7462 bp in length
 * 125344 125443: gap of 100 bp
 * 125444 132508: contig of 7065 bp in length
 * 132509 132608: gap of 100 bp
 * 132609 141190: contig of 8582 bp in length
 * 141191 141290: gap of 100 bp
 * 141291 152556: contig of 11266 bp in length
 * 152557 152656: gap of 100 bp
 * 152657 163590: contig of 10934 bp in length
 * 163591 163691: gap of 100 bp
 * 163691 176619: contig of 13929 bp in length.

FEATURES

source

misc_feature
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11"
 /clone="RP11-689A10"
 /clone_id="RP11 Human Male BAC"
 1..1333
 /note="assembly-fragment"
 1434..2787
 /note="assembly-fragment"
 2888..3998
 /note="assembly-fragment"
 4099..5777
 /note="assembly-fragment"
 5878..7685
 /note="assembly-fragment"
 7786..8176
 /note="assembly-fragment"
 clone_end:sp6
 vector_side:right
 8277..9926
 /note="assembly-fragment"
 10027..11499
 /note="assembly-fragment"
 11600..13616
 /note="assembly-fragment"
 13717..16073
 /note="assembly-fragment"
 16174..18510
 /note="assembly-fragment"
 18611..20495
 /note="assembly-fragment"
 20596..23548
 /note="assembly-fragment"
 23649..25910
 /note="assembly-fragment"
 26011..28060
 /note="assembly-fragment"
 28161..30134
 /note="assembly-fragment"
 30235..32646
 /note="assembly-fragment"
 32747..35096
 /note="assembly-fragment"
 clone_end:17
 vector_side:right
 35197..37696
 /note="assembly-fragment"
 37797..40539
 /note="assembly-fragment"
 40640..45518
 misc_feature

CDS
 /gene="UGT1A7"
 /product="UDP glucuronosyltransferase 1A7"
 join(98552, .99406,181788, .181919,182603, .182690,
 182974, .183193,187016, .187313)
 /gene="UGT1A7"
 /codon_start=1
 /product="UDP glucuronosyltransferase 1A7"
 /protein_id="AAG30419.1"
 /db_xref="GI:1118744"
 /translation="MARCGWGLPLPYCLLLTCGFAKAGLIVPDGSHWFTQSV
 VEKLLRGHEVYVWPEVSMQSGRLNCVKTSTSTLEDDREWEVADARWAPL
 RSNPGLTSSNGIPLDFNSCRSLKDKKLVKEYESCPDPAVLDPAFCGLIYAKY
 FLSPVVFARGLTCHYIEEGACCPAPSYVPLGLGFDATPEKRYRNHIEHL
 FCPYFKNVLEIASLLOTPTVATLSTLRTDVEYKPKYPMNMTIGTIGT
 CHQKRPVPEFAYINAGSGHIVFSLGSMSEIPEKKAMADALGKIPOTVARY
 TGTSPNLANNTILYKWLPNNDLGHPRATITAGSHGYESSICNGPMPWML
 DQDKNAKREKAGAGTINLENTSEDLNAKAVINDKSKYENINLSLKHDPVE
 PDLAVFVVEFVRKKGAPHLRPAADLWYQYHSLDVLGFLAVLVAVTFECCCA
 YGRKCLQKKGKVKAKHRSKTH"
 109467, .187313
 /gene="UGT1A6"
 109467, .109473
 /gene="UGT1A6"
 join(109610, .110470,181788, .181919,182603, .182690,
 182974, .183193,187016, .187313)
 /gene="UGT1A6"
 /product="UDP glucuronosyltransferase 1A6"
 join(109610, .110470,181788, .181919,182603, .182690,
 182974, .183193,187016, .187313)
 /gene="UGT1A6"
 /codon_start=1
 /product="UDP glucuronosyltransferase 1A6"
 /protein_id="AAG30420.1"
 /db_xref="GI:1118745"
 /translation="MAGLRSRPSAGVFLALMGVVDKILVYPDGSHWFTQSV
 IYVLSDRGHEVYVWPEVSMQSGRLNCVKTSTSTLEDDREWEVADARWAPL
 RSNPGLTSSNGIPLDFNSCRSLKDKKLVKEYESCPDPAVLDPAFCGLIYAKY
 FLSPVVFARGLTCHYIEEGACCPAPSYVPLGLGFDATPEKRYRNHIEHL
 FCPYFKNVLEIASLLOTPTVATLSTLRTDVEYKPKYPMNMTIGTIGT
 CHQKRPVPEFAYINAGSGHIVFSLGSMSEIPEKKAMADALGKIPOTVARY
 TGTSPNLANNTILYKWLPNNDLGHPRATITAGSHGYESSICNGPMPWML
 DQDKNAKREKAGAGTINLENTSEDLNAKAVINDKSKYENINLSLKHDPVE
 PDLAVFVVEFVRKKGAPHLRPAADLWYQYHSLDVLGFLAVLVAVTFECCCA
 YGRKCLQKKGKVKAKHRSKTH"
 129525, .187313
 /gene="UGT1A5"
 129525, .129531
 /gene="UGT1A5"
 join(129600, .130466,181788, .181919,182603, .182690,
 182974, .183193,187016, .187313)
 /gene="UGT1A5"
 /product="UDP glucuronosyltransferase 1A5"
 join(129600, .130466,181788, .181919,182603, .182690,
 182974, .183193,187016, .187313)
 /gene="UGT1A5"
 /codon_start=1
 /product="UDP glucuronosyltransferase 1A5"
 /protein_id="AAG30421.1"
 /db_xref="GI:1118746"
 /translation="MATGLOVPLPOLANGLLLSVOPAESGKLYVPTDGSNISM
 REALRDHARGHOYVLTLEVMYIKTEENFTTTTALISWTODEDRLLGHTOSFE
 TEHLKMFRRKAIKNNNSLLIHRSCEVLLHNLALRHIAISDVVLTDPFLICAAV
 LAKTLSPAVFLRNPICDIDFGKQCPNPSYIPLLITNSDHTFLOAKMELVPL
 ALSTLCVAPAPYASLASELFQREVSVDYASVWLEFGDFMDYPRPIMNMTI
 GGINANGKPLSGEFAVINAGSGHIVFSLGSMSEIPEKKAMADALGKIPOTV
 LMYRTGRSPNLANNTILYKWLPNNDLGHPRATITAGSHGYESSICNGPMPWML
 PLRGDQNAKREKAGAGTINLENTSEDLNAKAVINDKSKYENINLSLKHDPVE
 PDLAVFVVEFVRKKGAPHLRPAADLWYQYHSLDVLGFLAVLVAVTFECCCA
 YGRKCLQKKGKVKAKHRSKTH"
 Query Match 100.0%; Score 20; DB 89; Length 198872;
 Best Local Similarity 100.0%; Pred. No. 0.48; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtcacgtgacacagtcacac 20
 Db 174939 gtcacgtgacacagtcacac 174938
 RESULT 9
 AF135466
 LOCUS AF135466 200 bp DNA PRI 21-NOV-1999
 DEFINITION Pongo pygmaeus upe-glucuronosyltransferase 1A1 (UGT1A1) gene,
 promoter region and partial cds.
 ACCESSION AF135466
 VERSION AF135466.1 GI:6456549
 KEYWORDS
 SOURCE orangutan,
 Pongo pygmaeus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pongo.
 REFERENCE 1 (bases 1 to 200)
 AUTHORS Hall,D., Ybazaeta,G., Destro-Bisoli,G., Petzl-Erler,M.L. and DI
 Rienzo,A.
 TITLE Variability at the uridine diphosphate glucuronosyltransferase 1A1
 promoter in human populations and primates
 JOURNAL Pharmacogenetics (1999) in press
 REFERENCE 2 (bases 1 to 200)
 AUTHORS Ybazaeta,G., Hall,D. and DI Rienzo,A.
 TITLE Direct Substitution
 JOURNAL Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
 E. 57th street, Chicago, IL 60637, USA
 FEATURES
 source
 1. 200
 /organism="Pongo pygmaeus"
 /db_xref="taxon:9600"
 <121. >200
 /gene="UGT1A1"
 /product="UDP-glucuronosyltransferase 1A1"
 <121. >200
 /gene="UGT1A1"
 121. >200
 /gene="UGT1A1"
 /codon_start=1
 /product="UDP-glucuronosyltransferase 1A1"
 /protein_id="AAFO9177.1"
 /db_xref="GI:6456550"
 /translation="MAVESQGRPLVGLLLCYLGPVYSHA"
 BASE COUNT 40 a 46 c 62 g 52 t
 ORIGIN
 Query Match 92.0%; Score 18.4; DB 89; Length 200;
 Best Local Similarity 95.0%; Pred. No. 5.6;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 gtcacgtgacacagtcacac 20
 Db 23 gtcacgtgacacagtcacac 42
 RESULT 10
 AF135467
 LOCUS AF135467 208 bp DNA PRI 21-NOV-1999
 DEFINITION Trachypithecus obscurus upe-glucuronosyltransferase 1A1 (UGT1A1)
 gene, promoter region and partial cds.
 ACCESSION AF135467
 VERSION AF135467.1 GI:6456551
 KEYWORDS
 SOURCE dusky leaf monkey.
 ORGANISM Trachypithecus obscurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 Colobinae; Trachypithecus.
 REFERENCE 1 (bases 1 to 208)
 AUTHORS Hall,D., Ybazaeta,G., Destro-Bisoli,G., Petzl-Erler,M.L. and DI

TITLE
Rienzo, A.
Variability at the uridine diphosphate glucuronosyltransferase 1A1 promoter in human populations and primates

JOURNAL
Pharmacogenetics (1999) In press

REFERENCE
2 (bases 1 to 208)

AUTHORS
Ybazaeta, G., Hall, D. and Di Rienzo, A.

TITLE
Direct Submission

JOURNAL
Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924 E. 57th Street, Chicago, IL 60637, USA

FEATURES
SOURCE
1. 208
/organism="Trachypithecus obscurus"
/db_xref="taxon:54181"
/gene="UGT1A1"
/product="UDP-glucuronosyltransferase 1A1"
/gene="UGT1A1"
125..>208
/codon_start=1
/product="UDP-glucuronosyltransferase 1A1"
/protein_id="AA09178.1"
/db_xref="GI:6456552"
/translation="WVESODRPLVGLLCLVLPVYCHAG"

CDS
45 a 46 c 60 g 57 t

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 92.0%; Score 18.4; DB 89; Length 208;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY
1 gtcacgtgacacagtcacaac 20
|||||

DB
30 CTCACGTGACACATCAAC 49

RESULT 11
AF135462 234 bp DNA PRI 21-NOV-1999
LOCUS
DEFINITION
Pan paniscus UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
promoter region and partial cds.
ACCESSION
AF135462
AF135462.1 GI:6456541
KEYWORDS
pygmy chimpanzee.
SOURCE
Pan paniscus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
REFERENCE
1 (bases 1 to 234)
Hall, D., Ybazaeta, G., Destro-Bisol, G., Petzl-Erler, M.L. and Di Rienzo, A.
Variability at the uridine diphosphate glucuronosyltransferase 1A1 promoter in human populations and primates
Pharmacogenetics (1999) In press
2 (bases 1 to 234)
Ybazaeta, G., Hall, D. and Di Rienzo, A.
Direct Submission
Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924 E. 57th Street, Chicago, IL 60637, USA

FEATURES
SOURCE
1. 234
/organism="Pan paniscus"
/db_xref="taxon:9597"
/gene="UGT1A1"
/product="UDP-glucuronosyltransferase 1A1"
/gene="UGT1A1"
148..>234
/codon_start=1

gene
CDS
148..>234
/gene="UGT1A1"
/product="UDP-glucuronosyltransferase 1A1"
/codon_start=1

/product="UDP-glucuronosyltransferase 1A1"
/protein_id="AA09173.1"
/db_xref="GI:6456542"
/translation="WVESODRPLVGLLCLVLPVYCHAG"

BASE COUNT
ORIGIN
50 a 53 c 69 g 62 t

Query Match
Best Local Similarity 92.0%; Score 18.4; DB 89; Length 234;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY
1 gtcacgtgacacagtcacaac 20
|||||

DB
46 CTCACGTGACACATCAAC 65

RESULT 12
AF135463 234 bp DNA PRI 21-NOV-1999
LOCUS
DEFINITION
Pan troglodytes UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
promoter region and partial cds.
ACCESSION
AF135463
AF135463.1 GI:6456543
KEYWORDS
chimpanzee.
SOURCE
Pan troglodytes
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
REFERENCE
1 (bases 1 to 234)
Hall, D., Ybazaeta, G., Destro-Bisol, G., Petzl-Erler, M.L. and Di Rienzo, A.
Variability at the uridine diphosphate glucuronosyltransferase 1A1 promoter in human populations and primates
Pharmacogenetics (1999) In press
2 (bases 1 to 234)
Ybazaeta, G., Hall, D. and Di Rienzo, A.
Direct Submission
Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924 E. 57th Street, Chicago, IL 60637, USA

FEATURES
SOURCE
1. 234
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/gene="UGT1A1"
/product="UDP-glucuronosyltransferase 1A1"
/gene="UGT1A1"
148..>234
/codon_start=1
/product="UDP-glucuronosyltransferase 1A1"
/protein_id="AA09174.1"
/db_xref="GI:6456544"
/translation="WVESODRPLVGLLCLVLPVYCHAG"

BASE COUNT
ORIGIN
50 a 53 c 69 g 62 t

Query Match
Best Local Similarity 92.0%; Score 18.4; DB 89; Length 234;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY
1 gtcacgtgacacagtcacaac 20
|||||

DB
46 CTCACGTGACACATCAAC 65

RESULT 13
MUSPC326A 4201 bp mRNA ROD 27-APR-1993
LOCUS
DEFINITION
Mus musculus protein P0326 mRNA, complete cds.

ACCESSION M95564
VERSION M95564.1 GI:200240
KEYWORDS protein PC326
SOURCE Mus musculus (strain BALB/c, sub-species domesticus) plasmacytoma CDNA to mRNA.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 4201)
AUTHORS Eckhardt, L.A., Laskov, R., Kuehl, W.M., Bergsagel, P.L. and Timplin, C.R.
TITLE Sequence and expression of a murine CDNA encoding PC326, a novel gene expressed in plasmacytomas but not normal plasma cells
JOURNAL Oncogene 7, 2059-2064 (1992)
MEDLINE 93026383
FEATURES

source 1..4201
location/qualifiers
organism "Mus musculus"
strain "BALB/c"
sub-species "domesticus"
db_xref "taxon:10090"
cell_line "4N2.1"
cell_type "plasma cell"
tissue_type "plasma cytoma"
repeat_region <1..1400
repeat "B1 repeat"
gene 1514..4201
gene "protein PC326"
CDS 1514..3757
note "protein PC326"
note "member of beta transducin repeat 'mosaic' protein family; putative"
codon_start 1
product "protein PC326"
protein_id "AA339895.1"
db_xref "GI:200241"
translation "MSHSESTYMAETAPENISILSGIGTSGALVDTKTSIDKTP RVSLTSSDVTGTEDSVLTQSTDVNSVDSYQEGEDDDDEDDKDDSL EDSNFIISCLNENYIPQWENGEVVEEQSGRPHPELEAGEVVEGGGSGSLFVFE LEAGVEAONVONLFHRELEGEVVEAQNVSMPPEYELAGVYAEVYQFOR YELAREVIGAGCGGSLRHVGLSGGEVYATVRRLLIHHELEGGEDVDQESSEM HEETSDSSROYDIEDSDIDEMTALFETSPPRNNVLSALRDQGLSSGRVYAC GARELVORELEHVEHSGCVNTVHNQGTLLASGSDLVIVYMWLKKRVLNFD SGHNNILQAKFLPNCNDAILAMCGRGQVAVQLSAVAGTMTKLVKKGASHRLG LEPSPREFLTSGEDAVFENIDLQAHAPSKLLVKGDKKGLVYFVNPVAVYQFA VGGODQFMRIYDQKIDENVANNGLKKECPHLLSSDYPAILFTSLMVSQGTILLAS NDEDIYIFNNSDSDAQAAYKRGKRRNNSVKGVTYPRRSFTVSSGDCGHFTKX SSCQIVFLEADGGTINCIDSHPTLYLVASSGLDEVRINSPIAEBSKRLAQLKVI KINKAKDNFTLRHTSLFNNSMLCFILMSHTVTSNGRSMGRIRINAGGDFSDSSSSS EETNQES"
repeat_region 1882..2304
standard_name "PC326"
note "20 amino acid repeat; putative"
tbl_type "tandem"
3'UTR 3575..4201
gene "protein PC326"
note "putative"
polya_signal 4168..4174
gene "protein PC326"
note "putative"
polya_site 4186
gene "protein PC326"
note "putative"
BASE COUNT 1331 a 755 c 1006 g 1101 t 8 others
ORIGIN

Query Match 87.0%; Score 17.4; DB 94; Length 4201;
Best Local Similarity 94.7%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OR 1 gtacagtgcacagcaaa 19
||||| |||||||||

Db 3642 GTCATGTCACAGTCAAA 3660

RESULT 14
LOCUS AF069716/c
DEFINITION Arabidopsis thaliana BAC F644, Chromosome IV, near 60.5 cM,
ACCESSION AF069716
VERSION AF069716.1 GI:212102
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
AUTHORS Parnell, L.D., Huang, E.N., de la Bastide, M., Calma, C., Habermann, K., Schuttz, K., Gnoj, L., Dedhia, N.N. and Simon, N.
TITLE Genomic sequence of Arabidopsis thaliana BAC F644, Chromosome IV, near 60.5 cM
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99123)
AUTHORS Parnell, L.D., Huang, E.N., de la Bastide, M., Calma, C., Habermann, K., Schuttz, K., Gnoj, L., Dedhia, N.N. and Simon, N.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-1998) Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

COMMENT BAC F644 was sequenced as part of the arabidopsis genome sequencing effort of the Cold Spring Harbor Consortium. For additional information, please see <http://www.cshl.org/arabweb>. F644 overlaps with T1143 towards the centromere. Fingerprint data place F644 about 150 kbp to the left of the KNT3 marker.
FEATURES

source 1..99123
location/qualifiers
organism "Arabidopsis thaliana"
cultivar "Columbia"
db_xref "taxon:3702"
chromosome "v"
map "60.5 cM"
clone "F644"
note "hybridizing YAC is C102E3"
BASE COUNT 31735 a 16898 c 17599 g 32691 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 12; Length 99123;
Best Local Similarity 94.7%; Pred. No. 18;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OR 2 tcaagtgcacagcaaac 20
||||| |||||||||
Db 49925 TCACGTACACAGTCAAC 49907

RESULT 15
LOCUS AF135470
DEFINITION Arabidopsis thaliana BAC F644, Chromosome IV, near 60.5 cM,
ACCESSION AF135470
VERSION AF135470.1 GI:6456557
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
AUTHORS Parnell, L.D., Huang, E.N., de la Bastide, M., Calma, C., Habermann, K., Schuttz, K., Gnoj, L., Dedhia, N.N. and Simon, N.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-1998) Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA
COMMENT BAC F644 was sequenced as part of the arabidopsis genome sequencing effort of the Cold Spring Harbor Consortium. For additional information, please see <http://www.cshl.org/arabweb>. F644 overlaps with T1143 towards the centromere. Fingerprint data place F644 about 150 kbp to the left of the KNT3 marker.
FEATURES

JOURNAL Pharmacogenetics (1999) In press
 REFERENCE 2 (bases 1 to 207)
 AUTHORS Ybazaeta,G., Hall,D. and Di Rienzo,A.
 TITLE Direct Submission
 JOURNAL Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
 E. 57th Street, Chicago, IL 60637, USA
 FEATURES
 SOURCE Location/Qualifiers
 1..207

organism="Cebus apella"
 /db_xref="taxon:9515"
 <144..>207

gene
 /gene="UGT1A1"
 <144..>207
 /product="UDP-glucuronosyltransferase 1A1"

CDS
 /gene="UGT1A1"
 144..>207
 /codon_start=1
 /product="UDP-glucuronosyltransferase 1A1"
 /protein_id="AA09181.1"
 /db_xref="GI:6456558"
 /translation="MPACPGPAVCAAGPGSVPCWE"

BASE COUNT 43 a 46 c 62 g 56 t
 ORIGIN

Query Match 84.0%; Score 16.8; DB 89; Length 207;
 Best Local Similarity 90.0%; Pred. No. 52;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtacgtgacacagtcacac 20
 |||||
 Db 23 gtctctgacacatcattac 42

Search completed: July 25, 2001, 05:17:09
 Job time: 9232 sec